V. DIRECT EMPIRICAL FINDINGS ON SITE CHOICE AND WATER QUALITY PERCEPTION

In this chapter two types of analysis are used to detect the response of recreationists to water quality. First respondents were asked to rank water quality along with other determinants of site choice. In general, this approach finds that proximity and beach characteristics (facilities, cleanliness and setting) are much more important than water quality in determining site choice. If water quality improvements open sites close to major population centers, then benefits may be generated.

Second, the relationship between objective measures of water quality and the subjective water quality rating is probed in Section V.2. Logic suggests that strong correlation between objective and subjective measures is a necessary but not sufficient condition for demand to show any response to changes in water quality. Despite a rigorous analysis, of the data, we find weak, if any, association between the objective and subjective measures. While the engineer or public health scientist may measure improvements or declines in water quality, the public will not, it seems, perceive those changes.

1. Direct Questioning

Respondents were questioned directly concerning importance of various factors, including water quality to their recreational behavior. Four questions were posed:

1.1 The Favorite Site

Let's talk about the beach, lake, or river site you visited most. That was _______ site number ________. (Hand Respondent Card D)

A. Why do you visit this site most often? (Code most important reasons)

a. it is close
b. it is cheap
c. the water temperature is nice
d. the water quality is good
e. my family always came here
f. not too crowded
g. nice setting
h. beach is clean
l. nice facilities
j. my friends go there
k. other ______________

This of all the questions is probably the best indicator of behavior because the respondent considers and explains specific rather than generic behavior. Responses to this question are shown in Table V-1. Proximity is clearly the most important factor (47.5%). That friends go there, what we describe as a cultural factor, is the second most important reason (12.3%). Factors related to the beach quality (lack of litter--10.3%, and setting--11.7%) are the third and fourth most frequently mentioned responses, but are much less important than proximity. Water quality only gains 3.9% of the responses.

Response was tested against income, family size, education, occupation, race, amount of recreational equipment, and the amount of leisure time, automobile ownership, use of public transit and vacation time. Only income and family size affected the response distribution at a 5% level of significance. For all family sizes, proximity is the most important reason cited. The presence of friends is more important to larger families than

Table V-1

Reason for Choosing Favorite Site

Response		Number	Percentage
a.	it is close	170	47.5
b.	it is cheap	2	.6
c.	the water temperature is nice	11	3.1
d.	the water quality is good	14	3.9
e.	my family always came here	17	4.7
f.	not too crowded	13	3.6
g.	nice setting	42	11.7
h.	beach is clean	37	10.3
i.	nice facilities	8	2.2
j.	my friends go there	44	12.3

smaller ones. Similarly, larger families respond to water quality more readily than do smaller ones. These results are shown in Table V-2.

Table V-3 shows the income cross tabulation. Again proximity is always the most important reason, but declines in importance with higher incomes. Conversely, the importance of beach cleanliness increases moderately with higher incomes. The cell counts for water quality are too small to discern with any confidence the income trend, however.

Table V-2

Cross Tabulation of Reason for Visiting

Favorite Site and Family Size

REASON:	Family Size (# of members) 1	2	3	4	5	6	7	8	9	10
a. it is close	(5) 3.0 35.7	(27) 16.0 42.4	(22) 13.0 43.1	(52) 30.8 58.4	(26) 15.4 43.4	(16) 9.5 43.2	(8) 4.7 50.0	(8) 4.7 57.1	(2) 1.2 50.0	(3) 1.8 37.8
b. it is cheap	(0) 0 0	(1) 50.0 1.6	(0) 0 0	(0) 0 0	(0) 0 0	(1) 50.0 2.7	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0
c. the water temperature is nice	(0) 0 0	(1) 9.1 1.6	(0) 0 0	(4) 36.4 4.5	(1) 9.1 1.7	(2) 18.2 5.4	(1) 9.1 6.3	(0) 0 0	(1) 9.1 25.0	(1) 9.1 12.5
d. the water quality is good	(0) 0 0	(3) 21.4 4.7	(1) 7.1 2.0	(3) 21.4 3.4	(5) 35.7 8.3	(0) 0 0	(2) 14.3 12.5	(0) 0 0	(0) 0 0	(0) 0 0
e. my family always came here	(2) 11.8 14.3	(0) 0 0	(2) 11.8 3.9	(3) 17.6 3.4	(5) 29.4 8.3	(1) 5.9 2.7	(2) 11.8 12.5	(0) 0 0	(1) 5.9 25.0	(1) 5.9 12.5
f. not too crowded	(1) 7.7 7.1	(2) 15.4 3.1	(0) 0 0	(1) 7.7 1.1	(2) 15.4 3.3	(5) 38.5 13.5	(0) 0 0	(0) 0 0	(0) 0 0	(2) 15.4 25.0
g. nice setting	(2) 4.8 14.3	(8) 19.0 12.5	(9) 21.4 17.6	(12) 28.6 13.5	(2) 4.8 3.3	(5) 11.9 13.5	(2) 4.8 12.5	(2) 4.8 14.3	(0) 0 0	(0) 0 0
h. beach is clean	(2) 5.4 14.3	(13) 35.1 20.0	(8) 21.6 15.7	(6) 16.2 6.7	(6) 16.2 10.0	(2) 5.4 5.4	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0
i. nice facilities	(0) 0 0	(0) 0 0	(2) 25.0 3.9	(2) 25.0 2.2	(4) 50.0 6.7	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0
j. my friends go there	(2) 4.5 14.3	(9) 20.5 14.1	(7) 15.9 13.7	(6) 13.6 6.7	(9) 20.5 15.0	(5) 11.4 13.5	(1) 2.3 6.3	(4) 9.1 28.6	(0) 0 0	(1) 2.3 12.5

*Significant at 5% level, cells show number in ( ), row percentage and column percentages.

Table V-3

Cross Tabulation of Reasons*

For Choosing Favorite Site and Income

REASON:	Income Class 1	2	3	4	5	6	7	8	9	10	11
a. it is close	(8) 8.9 42.9	(15) 11.1 53.6	(20) 14.8 52.8	(19) 14.1 52.8	(16) 11.9 41.0	(31) 23.0 56.4	(12) 8.9 44.4	(4) 3.0 28.6	(3) 2.2 42.9	(0) 0 0	(3) 2.2 37.5
b it is cheap	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(1) 50.0 3.7	(1) 50.0 7.1	(0) 0 0	(0) 0 0	(0) 0 0
c. the water temperature is nice	(1) 11.1 3.6	(0) 0 0	(2) 22.2 4.3	(2) 22.2 5.6	(0) 0 0	(2) 22.2 3.6	(0) 0 0	(1) 11.1 7.1	(0) 0 0	(0) 0 0	(1) 11.1 12.5
d. the water quality is good	(2) 18.2 7.1	(1) 9.1 3.6	(2) 18.2 4.3	(0) 0 0	(2) 18.2 5.1	(4) 36.4 7.3	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0
e. my family always	(1) 7.1 .7	(0) 00 00	(4) 28.6 8.7	(1) 7.1 2.8	(3) 21.4 7.7	(1) 7.1 1.8	(2) 14.3 7.4	(1) 7.1 7.1	(0) 0 0	(0) 0 0	(1) 8.3 12.5
f. not too crowded	(0) 0 0	(2) 16.7 7.1	(1) 8.3 2.2	(0) 0 0	(3) 25.0 7.7	(1) 8.3 1.8	(1) 8.3 3.7	(2) 16.7 14.3	(0) 0 0	(1) 8.3 100.0	(1) 8.3 12.5
g. nice setting	(4) 11.1 14.3	(0) 0 0	(8) 22.2 17.4	(4) 11.1 11.1	(4) 11.1 10.3	(7) 19.4 12.7	(2) 5.6 7.4	(2) 5.6 14.3	(3) 8.3 42.9	(0) 0 0	(2) 5.6 25.0
h. beach is clean	(2) 6.1 7.1	(2) 6.1 7.1	(5) 15.2 10.9	(8) 24.2 22.2	(5) 15.2 12.8	(5) 15.2 9.1	(4) 12.1 14.8	(1) 3.0 7.1	(1) 3.0 14.3	(0) 0 0	(0) 0 0
i. nice facilities	(3) 42.9 10.7	(0) 0 0	(1) 14.3 2.2	(0) 0 0	(1) 14.3 2.6	(1) 14.3 1.8	(1) 14.3 3.7	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0
j. my friends go there	(3) 10.0 10.7	(8) 26.7 28.7	(3) 10.0 6.5	(2) 6.7 5.6	(5) 16.7 12.8	(3) 10.0 5.5	(4) 13.3 14.8	(2) 6.7 14.3	(0) 0 0	(0) 0 0	(0) 0 0

*Significant at 5% level, cells show number in ( ), row percentages and column percentages.

1.2 Characteristics Important for Site Choice

In choosing a site what are the three most important characteristics?

a. presence of a bathhouse/changing room
b. absence of litter
c. presence of a lifeguard
d. presence of a marine/boat launching facility
e. stocked game fish/good fishing
f. a natural setting
g. water temperature
h. water appearance
i. presence of other beach facilities
j. cost (parking fees, entry fees)
k. proximity
l. where your friends go
m. where your family always went
n. other ___________

We anticipated this question would yield less reliable results than the first one since it is more vague and general. Table V-6 shows the response to this question. Here absence of litter is the most important reason followed by the presence of beach facilities (bathhouse, lifeguard) and a nice setting, water appearance, rates, fifth, and proximity, sixth.

Several features of this response pattern are notable. The most obvious is the relative lack of importance ascribed to proximity. Two explanations suggest themselves. First, when considering the generic question of motivation, respondents discount proximity, although it is quite important to determine actual behavior. An alternative hypothesis is that many more respondents understood the meaning of "it is close" than knew the definition of "proximity."

The responses were tested against income, family size, race, occupation, education, amount of recreational equipment, automobile ownership, amount of leisure time each week, vacation time and use of public transit.

Table V-4

Important Characteristics for Site Choice

Characteristic	Most Important #	Most Important %	2nd Most Important #	2nd Most Important %	3rd Most Important #	3rd Most Important %
a. presence of a bathhouse/ changing rooms	62	13.7	28	6.3	33	7.6
b. absence of litter	141	31.1	109	24.5	48	11.0
c. presence of a lifeguard	49	10.8	48	10.8	37	8.5
d. presence of a marina/ boat launching facility	5	1.1	12	2.7	6	1.4
e. stocked game fish/ good fishing	5	1.1	5	1.1	10	2.3
f. a natural setting	52	11.5	37	8.3	42	9.6
g. water temperature	14	3.1	38	8.6	26	5.9
h. water appearance	43	9.5	71	16.0	74	16.9
i. presence of other beach facilities	4	.9	16	3.6	16	3.7
j. cost (parking fees, entry fees)	3	.7	31	7.0	43	9.8
k. proximity	37	8.2	17	3.8	44	10.1
l. where your friends go	18	4.0	16	3.6	25	5.7
m. where your family always went	7	1.5	5	1.1	14	3.2
n. other	13	2.9	11	2.5	19	4.3

The null hypothesis of independent classification can be rejected at the 5% level for education and occupation. The contingency tables are presented in Tables V-5 and V-6, respectively. Higher levels of education lead to a greater sensitivity to a natural setting. At the same time, proximity becomes more important with increased education. Because setting and proximity are inversely related, this table suggests that respondents not understanding the definition of "proximity" may explain, at least in part, the markedly differing results from these two questions.

These results have two interesting implications, one methodological and one substantive. The first is that the wording of the questionnaire is of great importance to subsequent findings. Although our survey instrument was carefully developed, reviewed and pretested, this anomaly persisted and seems to have made a difference.

Secondly, facilities appear to be important to recreation demand. Any recreation benefits from water quality improvements may not be obtained unless further investments in beaches, changing facilities, maintenance and lifeguards are made. Additional money, perhaps raised through user fees, would be required to provide these facilities.

Table V-5

Most Important Site Characteristics
Tabulated by Education

CHARACTERISTIC	Education 1	2	3	4	5	6	7
a. presence of a bath-house/changing rooms	(0) 0 0	(1) 1.6 7.7	(21) 34.4 17.9	(13) 21.3 15.9	(5) 8.2 17.9	(17) 27.9 14.2	(4) 6.6 4.8
b. absence of litter	(1) .7 16.7	(5) 3.6 38.5	(31) 22.3 26.5	(33) 23.7 40.2	(8) 5.8 28.6	(36) 25.9 30.0	(25) 18.0 30.1
c. presence of life-guard	(2) 4.1 33.3	(2) 4.1 15.4	(15) 30.6 12.8	(9) 18.4 11.0	(3) 6.1 10.7	(9) 18.4 7.5	(9) 18.4 10.8
d. presence of a marina/boat launching facility	(0) 0 0	(0) 0 0	(1) 20.0 .9	(1) 20.0 1.2	(0) 0 0	(3) 60.0 2.5	(0) 0 0
e. stocked gems fish/good fishing	(0) 0 0	(0) 0 0	(3) 60.0 2.6	(0) 0 0	(0) 0 0	(1) 20.0 .8	(1) 20.0 1.2
f. a natural setting	(1) 1.9 16.7	(1) 1.9 7.1	(7) 13.5 6.0	(11) 21.2 13.4	(2) 3.8 7.1	(16) 30.8 13.3	(14) 26.9 16.9
g. water temperature	(0) 0 0	(0) 0 0	(7) 50.0 6.0	(0) 0 0	(2) 14.3 7.1	(4) 28.6 3.3	(1) 7.1 1.2
h. water appearance	(1) 2.3 16.7	(0) 0 0	(11) 25.6 9.4	(8) 18.6 9.8	(2) 4.7 7.1	(11) 25.6 9.2	(10) 23.3 12.0
i. presence of other beach facilities	(1) 25.0 16.7	(0) 0 0	(1) 25.0 .9	(0) 0 0	(1) 25.0 3.6	(1) 25.0 .8	(0) 0 0
j. cost (parking fees, entry fees)	(0) 0 0	(0) 0 0	(1) 33.3 .9	(1) 33.3 1.2	(0) 0 0	(0) 0 0	(1) 33.3 1.2
k. proximity	(0) 0 0	(2) 5.4 15.4	(7) 18.9 6.0	(2) 5.4 2.4	(3) 8.1 10.7	(11) 29.7 9.2	(12) 32.4 14.5
l. where your friends go	(0) 0 0	(1) 5.6 7.7	(8) 44.4 6.8	(3) 16.7 3.7	(0) 0 0	(6) 33.3 5.0	(0) 0 0
m. where your family always went	(0) 0 0	(1) 16.7 7.7	(1) 16.7 .9	(0) 0 0	(2) 33.3 7.1	(1) 16.7 .8	(1) 16.7 1.2
n. other	(0) 0 0	(0) 0 0	(3) 23.1 2.6	(1) 7.7 1.2	(0) 0 0	(4) 30.8 3.3	(5) 38.5 6.0

Table shows cell Count in ( ), row percentages and column percentages.

Table V-6

Most Important Site Characteristics Tabulated by Occupation

Characteristic	Occupation 1	2	3	4	5	6	7	8	9	10
a. presence of a bathhouse/changing rooms	(15) 25.4 12.3	(7) 11.9 9.5	(2) 3.4 16.7	(13) 22.0 15.7	(4) 6.8 7.8	(1) 1.7 4.3	(9) 15.3 34.6	(1) 1.7 11.1	(1) 1.7 16.7	(6) 10.2 16.7
b. absence of litter	(37) 27.2 30.3	(21) 15.4 28.4	(5) 3.7 41.7	(27) 19.9 32.5	(15) 11.0 29.4	(10) 7.4 43.5	(6) 4.4 23.1	(4) 2.9 44.4	(2) 1.5 33.3	(9) 6.6 25.0
c. presence of lifeguard	(15) 31.3 12.3	(8) 16.7 10.8	(1) 2.1 8.3	(10) 20.8 12.0	(5) 10.4 9.8	(2) 4.2 8.7	(2) 4.2 7.7	(0) 0 0	(2) 4.2 33.3	(3) 6.3 8.3
d. presence of a marina/boat launching facility	(0) 0 0	(2) 40.0 2.7	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(3) 60.0 8.3
e. stocked game fish/good fishing	(1) 20.0 .8	(1) 20.0 1.4	(0) 0 0	(2) 40.0 2.4	(0) 0 0	(1) 20.0 4.3	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0
f. a natural setting	(18) 34.6 14.8	(7) 13.5 9.5	(0) 0 0	(9) 17.3 10.8	(9) 17.3 10.8	(3) 5.8 13.0	(3) 5.8 11.5	(1) 1.9 11.1	(1) 1.9 16.7	(1) 1.9 2.8
g. water temperature	(3) 23.1 2.5	(2) 15.4 2.7	(0) 0 0	(4) 30.8 4.8	(4) 30.8 4.8	(0) 0 0	(1) 7.7 3.8	(0) 0 0	(0) 0 0	(0) 0 0
h. water appearance	(10) 23.3 8.2	(13) 30.2 17.6	(0) 0 0	(4) 9.3 4.8	(6) 14.0 11.8	(2) 4.7 8.7	(1) 2.3 3.8	(3) 7.0 33.3	(0) 0 0	(4) 9.3 11.1
i. presence of other beach facilities	(2) 50.0 1.6	(0) 0 0	(2) 50.0 16.7	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0
j. cost (parking fees, entry fees)	(1) 33.3 .8	(0) 0 0	(0) 0 0	(0) 0 0	(1) 33.3 2.0	(0) 0 0	(0) 0 0	(0) 0 0	(0) 0 0	(1) 33.3 2.8
k. proximity	(13) 35.1 10.7	(7) 18.9 9.5	(0) 0 0	(6) 16.2 7.2	(4) 10.8 7.8	(3) 8.1 13.0	(0) 0 0	(0) 0 0	(0) 0 0	(4) 10.8 11.1
l. where your friends go	(2) 11.1 1.6	(3) 16.7 4.1	(1) 5.6 8.3	(5) 27.8 6.0	(2) 11.1 3.9	(0) 0 0	(2) 11.1 7.7	(0) 0 0	(0) 0 0	(3) 16.7 8.3
m. where your family always went	(1) 16.7 .8	(0) 0 0	(0) 0 0	(2) 33.3 2.4	(2) 33.3 3.9	(0) 0 0	(1) 16.7 3.8	(0) 0 0	(0) 0 0	(0) 0 0
n. other	(4) 30.8 3.3	(3) 23.1 4.1	(1) 7.7 8.3	(1) 7.7 1.2	(0) 0 0	(1) 7.7 4.3	(1) 7.7 3.8	(0) 0 0	(0) 0 0	(2) 15.4 5.6

Table shows cell count in ( ), row percentages and column percentages.

1.3 Not Visiting Closest Site

The third question asks the converse of the first one:

(If the respondent did not visit the closest site, ask:) (Hand respondent Card B)

_____________ beach is the major recreation site closest to your home, yet you did not mention having visited it. Here are some reasons, which one best explains why you did not visit that site?

a. not aware of that site
b. do not like the facilities
c. too crowded
d. beach too dirty
e. water too cold
f. water too dirty
g. don't own auto, not accessible by public transportation
h. too expensive
i. not interested in the activities available there
j. other (please specify) _______________

Here we control for proximity to assess the rationale behind site choice. The principal shortcoming of this question is that, since the respondent does not visit the closest site, his knowledge of it may be dated or secondhand.

Table V-2 shows the response distribution to this question. It is remarkable, given the apparent importance of proximity to attendance, that 60.2% of the respondents did not visit the closest sites. Of course, the second most close site was, in many sample clusters, quite close by. The importance of this finding is mitigated somewhat by the widespread ignorance of the closest site (response a). The ignorance hypothesis is further confirmed by the second most important reason, "not interested in the activities available there," because the beaches were offered quite homogenous activities: swimming, boating, fishing, picnicking, bicycling, strolling and informal sports were available at all, and only a few offer facilities for tennis, basketball and other similar specialized sports.

Table V-7

Distribution of Reasons for not Visiting Closest Site

Reason	No.	Percent
a. not aware of that site	69	24.6
b. do not like the facilities	14	5.0
c. too crowded	31	11.0
d. beach too dirty	24	8.5
e. water too cold	0	0
f. water too dirty	32	11.4
g. don't own auto, not accessible by public transportation	2	.7
h. too expensive	0	0
i. not interested in the activities. available there	39	13.9
j. other	70	24.9
TOTAL	281	

Dirty water and crowding ranked third and fourth, respectively as major deterents to attendance. The hypothesis that good water quality does not encourage attendance, but bad water quality discourages it suggests itself, but is not confirmed by the willingness-to-pay analysis presented below. Judging from the low correlations between water quality and water quality perceptions, "the water is too dirty," may be another way of saying "I don't visit the site because I am told it is not very nice." Hence a public agency might reduce attendance at a polluted site by identifying it as such. And, the converse may also be true: water quality improvements may not increase use unless there is adequate publicity that the beach is open for swimming or that the water quality has been improved. This may be particularly important for sites where water quality has been poor for some time, such as the lower Charles River in Boston.

The obvious hypotheses concerning the effects of income, race, education, occupation, automobile ownership, public transit usage, vacation time, and leisure time on reasons for selecting a site were tested via contingency tables and no effect was found to be statistically significant at the 5% level. Once those who do visit the closest site have been removed from the sample, it is easy to see why those remaining do not differ along these socioeconomic lines, but the income of those visiting the closest site is not statistically different from the income of those who visit more distant sites.

1.4 Importance of Various Water Characteristics

The final direct question used to probe the relationship between recreation behavior and water quality focused on the characteristics people feel are important to good water quality:

> Thinking of water quality, attractiveness of the water for swimming depends on the color, odor, clearness, amount of floating debris or scum, and the amount of aquatic weeds. Which characteristic

is the most important? 2nd most important? Please rank these characteristics.

a. color
b. odor
c. clearness
d. floating debris
e. aquatic weeds.

Responses to this question are tabulated in Table V-8. Clarity (the converse of turbidity) and the absence of floating debris appear to be the most important parameters of water quality. These results contrast with the observed ratings which show only color to be correlated with water quality perception (Section IV-4 above). In this ranking color is next to last in importance. Several explanations for this contrast are possible. The best is that this question, generic rather than specific, is not a reliable indicator of perception. Another is that because turbidity and color are intercorrelated (R^2=.72 for our sample of sites) the two were confused in this question. In other words, respondents did not understand the distinction between color and turbidity. In hindsight it may have hindered the analysis to include both.

The presence of aquatic weeds is of minor importance. This may be due to the low incidences of eutrophication found in Boston's cold weather climate.

Table V-8

Importance of Various Water Quality Characteristics

Characteristic	Most Important		2nd		3rd		4th		Least Important	
Characteristic	#	%	#	%	#	%	#	%	#	%
a. · color	39	8.4	73	15.6	78	16.7	145	31.0	110	23.6
b. odor	78	16.7	142	30.4	130	27.8	64	13.7	26	5.6
c. clearness	157	33.6	78	16.7	75	16.1	82	17.6	54	11.6
d. floating debris	157	33.6	115	24.6	80	17.1	56	12.0	28	6.0
e. aquatic weeds	15	3.2	38	8.1	76	16.3	88		217	46.5

1.5 Conclusions

In sum, the responses to these questions do not seem to support any hypothesis which relates recreation behavior to water quality. They suggest proximity is the most important determinant of a site choice. To the extent that improvements in water quality will open up beaches proximal to large numbers of people, the water quality improvement will lead to increased recreation benefits. This would be the case in many urban places, and particularly Boston.

A secondary conclusion is that recreation behavior is not overwhelmingly determined by socioeconomic variables. To a small extent higher levels of SES may reduce the sensitivity to distance and increase the propensity to visit the more distant, litter free beaches in a natural setting. Larger family size suggests a greater propensity to visit beaches where friends go.

Finally, the presence of facilities appears to be an important factor in site choice. If so, improvements in water quality should be accompanied by beach maintenance and capital investments to gain recreation benefits.

2. Public Perception of Water Quality

Do respondents agree on the quality of the water at individual sites? Does the public perception of water quality match the objective conditions? Which objective water qualith characteristics affect most strongly the respondent's perception of site conditions? These are the questions of this section.

The answers are the foundation for the demand models presented in Chapter VI. In particular, a link between perceived and objective water quality characteristics is a necessary but not sufficient condition to establish recreation benefits from water quality improvement.

2.1 Agreement Among Respondents

The first question presents the greatest analytical difficulties since there is at present no convenient methodology for assessing the degree of nominal scale agreement among multiple raters. There is a well-developed methodology for the case of two raters involving the kappa statistic but with more than two raters the only available approach appears to be to compute the full set of $\binom{n}{2}$ pairwise agreement statistics and to average them. This procedure can be applied when there is a small number of raters but it is manifestly impractical with several hundred raters.* Therefore, an informal analysis of the rating distribution must suffice. The distributions of water quality ratings for sites 1 to 29 are shown in Table V-9. With the exception of sites 6, 22 and 23, the distributions seem to be reasonably tight. Judging the degree of concensus by the percentage of total responses

*The problem of multiple raters is discussed in Fleiss [3] and Light [6]. Fleiss presents an application of the procedure described in the text to a case with six observers. This problem is also discussed briefly in Bishop *et al* [1].

Table V-9

Distribution of Ratings of Water Quality for 28 Sites

Site	# of Evaluations	% of Ratings in Category				
		1	2	3	4	5
1	24	12.5	29.2	29.2	16.7	12.5
2	44	15.9	34.1	31.8	15.9	2.3
3	98	9.2	17.3	33.7	27.6	12.2
4	119	37.0	29.4	22.7	8.4	2.5
5	10	20.0	20.0	30.0	20.0	10.0
6	13	23.1	15.4	15.4	38.5	7.7
7	14	42.9	28.6	7.1	14.3	7.1
8	27	25.9	44.4	18.5	3.7	7.4
9	7	14.3	28.6	42.9	14.3	0.0
10	9	22.2	33.3	33.3	0.0	11.1
11	11	18.2	18.2	54.5	9.1	0.0
12	12	16.7	33.3	41.7	8.3	0.0
13	13	30.8	38.5	23.1	7.7	0.0
14	5	20.0	20.0	40.0	20.0	0.0
15	41	34.1	29.3	26.8	9.8	0.0
16	124	4.8	12.1	28.2	30.6	24.2
17	57	0.0	3.5	12.3	40.4	43.9
18	86	3.5	3.5	11.6	45.3	36.0
19	45	6.7	13.3	26.7	35.6	17.8
20	28	3.6	3.6	21.4	21.4	50.0
21	18	0.0	5.6	16.7	38.9	38.9
22	34	26.5	14.7	23.5	14.7	20.6
23	23	26.1	34.8	21.7	0.0	17.4
24	18	50.0	16.7	22.2	5.6	5.6
25	24	8.3	16.7	29.2	33.3	12.5
26	46	4.3	17.4	23.9	34.8	19.6
27	8	37.5	12.5	12.5	25.0	12.5
28	20	20.0	15.0	30.0	20.0	15.0

NOTE: Rows sum to 100%, apart from rounding errors. The modal rating in each row is underlined. Site 29 was only rated by two respondents and is, therefore, omitted. (1=bad, 3=fair, 5=good) accounted for by the modal response, the concensus is somewhat greater, in general, for the sites with a higher modal water quality rating.

2.2 Accuracy of Perceptions

Given reasonably consistent ratings, the conceptually more important question of the accuracy of respondent's perceptions of water quality conditions can be considered. Before proceeding with this issue, recall that the yardstick for measuring the accuracy of public perceptions is the data obtained from our own water quality survey. Every effort was made to make these samples as representative as possible. With this caveat, consider Table V-10 which shows the correlation between water quality rating and the 16 objective measures of water quality. Negative correlations would be expected in all cases. With the exception of color, none of the correlations are statistically distinguishable from zero. The correlation between perceived water quality and color is only.moderate, equalling -.377. The low correlation might, of course, be due to the delay in implementing the survey.

To obtain more detailed evidence on the accuracy of the respondents' perceptions of water quality and, at the same time, in order to examine the relative importance of different water quality parameters in the formation of people's perceptions of water quality, we regressed the water quality ratings for all sites on various objective water quality variables. There are some statistical problems with this procedure arising from the special nature of the dependent variable. Firstly, the water quality rating (RWQUAL) is a discrete variable; respondents were asked to rate sites on the integer scale from 1 to 5. Because ordinary least squares regression does not constrain the predicted value of the dependent variable to be an integer, it is more difficult to assess the true degree of association between the dependent and independent variables on the basis of the filled regression equation. Secondly, it is possible to argue that RWQUAL is not a cardinal but an ordinal variable: a person who rates

Table V-10

Correlations Between Water Quality Rating and Water Quality Variables

Variable	Correlation[+]
OIL	-.1100
JTU	-.0796
COLOR	-.3777*
PH	.1032
ALK	.0953
TPOS	-.1553
NITR	-.1044
AMMO	-.1752
COD	-.0136
COLI	-.1340
TBAC	-.0606
TEMP	-.2550
FACTOR1	.1211
FACTOR2	-.0516
FACTOR 3	-.1986
FACTOR 4	-.0385

[+]All figures are based on 29 observations (sites), those with an asterisk are significant at the 5% level.

a site at 4 certainly likes it more than a site which he rates at 2, but not necessarily twice as much more. Ordinary least squares is not a desirable technique for handling this type of dependent variable. Rather, it is preferable to use the maximum likelihood estimation procedure which is described below.

We start, however, with some OLS regressions of RWQUAL on selected water quality variables and the composit water quality factors. The results of these regressions are shown in Table V-11. It is clear that the water quality ratings are significantly affected by all the water quality parameters, except OIL. The slope coefficients for most variables have the signs which we would expect; the only exceptions are the coefficients of squared pH deviations from a neutral value of 7, and of temperature. The sign of the coefficient for temperature may be an artifact of the sample since inner-harbor sites are both warmer and more polluted than the more distant ones. The performance of the factor scores as explanatory variables is somewhat disappointing: on the whole, they do not perform any better than the water quality parameters to which they are related. Factor 3, the clarity factor, performs best as would be expected. The bacterial factor, Factor 4, also has an adequate t-statistic. Among the most important parameters for explaining water quality ratings are TURBIDITY, COLOR, PHOSPHORUS, AMMONIA, and COLIFORM and TOTAL BACTERIA.* The explanatory power of the individual equations is low, but this is partly to be expected because of the discreteness of the dependent variable. We are thus led to the conclusion that, while there is a significant connection between objective water quality conditions and the subjective water quality ratings, the degree of association between them does not appear to be very great.

---

*The slope coefficients for TOTAL and COLIFORM BACTERIA appear somewhat similar and, indeed, when RWQUAL is regressed on both variables, the hypothesis that they have the same slope coefficient cannot be rejected.

Table V-11

Regression of Water Quality and Temperature Ratings on Water Quality Parameters

(984 observations)

RWQUAL = 3.057 + 0.00254 OIL		R^2=.000
(42.73) (0.36)		
RWQUAL = 3.256 - 0.0537 TURBIDITY		R^2=.024
(56.73) (4.88)		
RWQUAL = 3.41 - 0.0529 COLOR		R^2=.037
(48.57) (6.11)		
RWQUAL = 2.743 + 0.353 $(PH-7)^2$		R^2=.008
(22.08) (2.76)		
RWQUAL = 2.834 + 0.00263 ALKALINITY		R^2=.005
(24.44) (2.25)		
RWQUAL = 3.499 - 7.6351 PHOSPHORUS		R^2=.064
(53.2) (8.18)		
RWQUAL = 3.096 - 0.3117 NITROGEN		R^2=.006
(72.44) (2.45)		
RWQUAL = 3.287 - 0.4665 AMMONIA		R^2=.032
(59.17) (5.67)		
RWQUAL = 3.244 - 0.00534 COD		R^2=.007
(42.84) (2.64)		
RWQUAL = 3.165 - 0.0000542 COLIFORM BACTERIA		R^2=.022
(64.19) (4.66)		
RWQUAL = 3.215 - 0.0000164 TOTAL BACTERIA		R^2=.021
(62.32) (4.53)		
RWQUAL = 8.162 - 0.0773 TEMPERATURE		R^2=.025
(7.95) (4.96)		
RWQUAL = 3.037 + 0.0976 FACTOR 1		R^2=.004
(18.18) (2.02)		
RWQUAL = 3.059 - 0.0794 FACTOR 2		R^2= .002
(71.82) (1.55)		
RWQUAL = 2.964 - 0.2995 FACTOR 3		R^2=.024
(63.33) (4.89)		
RWQUAL = 3.054 - 0.1681 FACTOR 4		R^2=.013
(72.18) (3.58)		
RTEMP = 0.13 + 0.04082 TEMPERATURE		R^2=.007
(0.12) (2.57		

A subsidiary issue, which can conveniently be analyzed in the regression context, is the question of whether respondent's from households which participated in boating or fishing might have a different perception of water quality than other respondents. This could be tested by adding a dummy variable for participation in these activities to the regression in Table V-11 but this would not necessarily be the best procedure, since there is no presumption that fishers or boaters rate sites higher or lower than the public at large. Rather, the presumption is merely that they rate sites differently from other people. To test this hypothesis, we conducted separate regressions of RWQUAL on COLOR and COLI for respondents from households which participated in boating and/or fishing and for respondents from households which do not.* In addition, we conducted a regression on the full posted sample. The regression results are as follows:

FISHERS/BOATERS (551 Observations)

$$\underset{(34.2)}{\text{RWQUAL} = 3.353} - \underset{(2.54)}{0.033}\ \text{COLOR} - \underset{(2.54)}{0.0000449}\ \text{COLI} \qquad R^2 = .037,\ F = 10.65,\ SSR = 989.46$$

NON-FISHERS/BOATERS (429 Observations)

$$\underset{(35.54)}{\text{RWQUAL} = 3.501} - \underset{(4.45)}{0.06203}\ \text{COLOR} - \underset{(0.21)}{0.00000381}\ \text{COLI} \qquad R^2 = .06,\ F = 13.68,\ SSR = 636.98$$

FULL SAMPLE POOLED (980 Observations)

$$\underset{(48.91)}{\text{RWQUAL} = 3.415} - \underset{(4.70)}{0.0448}\ \text{COLOR} - \underset{(2.16)}{0.0000277}\ \text{COLI} \qquad R^2 = .043,\ F = 22.16,\ SSR = 1632.09$$

---

*These explanatory variables were chosen as being among the most important in the single variable regressions. Another variable which we attempted to include is PHOSPHORUS, but it turned out that this variable is highly collinear with COLOR and COLI, and, therefore, it was dropped from the regression.

Applying the standard Chow test for the equality of interceptor and slope coefficients, we find that the hypothesis of homogeneity between fishers/boaters and others cannot be rejected.

2.3 Ordinal Rankings Considered

A maximum likelihood estimation technique can explicitly allow for the fact that the dependent variable may provide only an ordinal raking of sites. The logic of the model is as follows. It is assumed that the respondent's true sentiment towards recreation sites, W, is a function of certain variables, X, and a random disturbance (representing, perhaps, random differences in tastes).

$$W_i = X_i\beta + V_i. \qquad \ldots \text{(1)}$$

The variable, W, is a continuous, cardinal measure of preference. However we do not observe it directly, instead we observe a discrete, ordinal variable, Y, which is a function of W and of certain "threshold" parameters, t_1, t_2, t_3, t_4.

$$\begin{aligned}
Y_i &= 1 \quad \text{if } W_i < t_1 \\
Y_i &= 2 \quad \text{if } t_i < W_i < t_2 \\
Y_i &= 3 \quad \text{if } t_2 < W_i < t_3 \\
Y_i &= 4 \quad \text{if } t_3 < W_i < t_4 \\
Y_i &= 5 \quad \text{if } W_i < t_4.
\end{aligned}$$

The threshold parameters together with the coefficient vector β are to be estimated from the observed data on Y and X.

The model represented by (1) and (2) is flexible, in that it specifically enables a test of the assumption that Y is cardinal: if the estimated t_j are (approximately) the integers from 1 to 4 we may conclude that Y is approximately a cardinal measure; in these circumstances, the results from the OLS regressions presented above would indeed be adequate. Otherwise, these conclusions would not be warranted. The model is also plausible in that it corresponds to the way in which one intuitively thinks of rating site conditions; it seems quite likely that people's underlying sentiments towards the sites are cardinal in nature but are then mapped into a discrete, ordinal variable in the process of answering the questionnaire.

In order to estimate the model it is necessary to make some assumptions about the distribution of the random variable u in (1). It is convenient to assume that these variables are independently and identically distributed, having a common normal distribution with mean zero and variance σ^2. The resulting likelihood function is:

$$\mathcal{L}(\beta t|X,Y) = \prod_{Y_i=1} P\left[\frac{t_1-X_i\beta}{\sigma}\right] \cdot \prod_{Y_i=2} \left\{P\left[\frac{t_2-X_i\beta}{\sigma}\right] - P\left[\frac{t_1-X_i\beta}{\sigma}\right]\right\} \cdot \ldots$$

$$\ldots \prod_{Y_i=4} \left\{P\left[\frac{t_4-X_i\beta}{\sigma}\right] - P\left[\frac{t_3-X_i\beta}{\sigma}\right]\right\} \cdot \prod_{Y_i=5} \left\{1-P\left[\frac{t_4-X_i\beta}{\sigma}\right]\right\}$$

where P[X] is the standard normal cumulative density function. In this model σ is not identifiable nor are all the threshold terms and the intercept in (1). As normalizations we take $\sigma=t_1=1$; with this assumption we can estimate both β and the differences $_j-t_{j-1})$ up to a multiplicative scale factor. The likelihood function is maximized by an iterative procedure which converged very rapidly in our experience.* Estimates of the variances and covariances of the coefficients are obtained from the Hessian of the likelihood function at the final iteration. From these estimates, the standard

*The convergence criterion criterion was that successive coefficient estimates must differ by less than .001 before the iteration stops. With our data this always happened by the sixth iteration.

t-test for significance can be derived since the computed test statistic asymptotically follows the t-distribution.

In order to implement the model, we focused on the relationship between the objective measures of color and coliform bacteria and subjective water quality relationships. The coefficient estimates are shown in the upper panel of Table V-12 (with the absolute value of the asymptotic t-statistic in parenthesis). It is noteworthy that the three bounded ranges are roughly (though not exactly) equally spaced, which tends to support the hypothesis that, at least in its middle range, RWQUAL is a cardinal measure. We can test the degree of association between the regressor variables and RWQUAL in at least two ways. The method is to compute the predicted scores using the estimated coefficients and see how many times the predicted score matches the actual score. The results of this test are very discouraging for the hypothesis of a strong correlation between objective site conditions and subjective perceptions: the predicted scores were all "1" ($\hat{W}$ ranged from -410 to -0.74), whereas only 155 of the 984 actual values of RWQUAL were 1. By this criterion, the model's fit is very poor. An alternative procedure to perform an analogue of the F-test in standard OLS regression to test the hypothesis is that the slope coefficients are jointly zero. For this purpose, we drop the regressor variables from the model while retaining the constant term and re-estimate the model. The resulting coefficient estimates are shown in Table V-12 in the lower panel. Although the likelihood function is lower for the second model than for the first, the difference is too small to be significant and hence we cannot reject the hypothesis that the slope coefficients are indeed zero.*

---

*An alternative measure of association would be the multiserial correlation coefficient between the predicted value of W and the actual value of RWQUAL. See Cox [2].

Table V-12

Maximum Likelihood Estimates of Ordinally Discrete Dependent Variable Model

W = perceived water quality

$$W = \underset{(32.96)}{2.293} - \underset{(4.52)}{0.0353}\ \text{COLOR} - \underset{(2.21)}{0.00002328}\ \text{COLI BACT}$$

$$\text{RWQUAL} = \begin{cases} 1 & \text{if } W < 1 \\ 2 & \text{if } 1 < W < \underset{(39.01)}{1.617} \\ 3 & \text{if } 1.617 < W < \underset{(43.73)}{2.262} \\ 4 & \text{if } 2.262 < W < \underset{(48.34)}{2.995} \\ 5 & \text{if } W < 2.995 \end{cases}$$

$$-\mathcal{L} = 1541.48$$

$$W = \underset{(41.46)}{2.002}$$

$$\text{RWQUAL} = \begin{cases} 1 & \text{if } W < 1 \\ 2 & \text{if } 1 < W < \underset{(39.44)}{1.605} \\ 3 & \text{if } 1.605 < W < \underset{(44.13)}{2.32} \\ 4 & \text{if } 2.32 < W < \underset{(48.59)}{2.947} \\ 5 & \text{if } W < 2.947 \end{cases}$$

$$-\mathcal{L} = 1562.48$$

2.4 Conclusions

In sum, the hypothesis that water quality perceptions are not linked to actual water quality cannot be rejected on the basis of our data. Aside from data problems described elsewhere in this report, the most obvious explanation of this result is that human sensory perception of water quality is inaccurate. This is not a surprising conclusion, particularly for the "invisible" contaminants such as bacteria, algal nutrients, COD, etc. Perhaps our only perception of water quality occurs when a beach is closed by the health department. Alternately, this result may derive from some undiscovered peculiarity of our sample of raters. In any case, this conclusion jeopardizes the search for a link between levels of water quality and demand.

CITED REFERENCES

1. Y.M. Bishop, S.E. Renberg & P.W. Holland, Discrete Multivariate Analysis, MIT Press, 1975.

2. N.R. Cox, "Estimation of the Correlation Between a Continuous a Discrete Variable, Biometrics (March 1974): 171-178.

3. J.L. Fleiss, "Measuring Nominal Scale Agreement Among Many Ratios," Psychological Bulletin (1971): 378-382.

4. Haggestrom, Notes on Discriminant Analysis, Logistic Regression, Rand Memorandum dated 4/3/74.

5. Yoel Haitovsky, Regression Estimation from Grouped Observations, 1973.

6. Richard J. Light, "Measures of Response Agreement for Qualitative Analysis," Psychological Bulletin (1971): 365-377.

VI. WILLINGNESS-TO-PAY

The willingness-to-pay survey method is frequently used for determining the value of public goods. This method, in essence, directly constructs a demand curve and its concomitant consumer surplus integral. Davis [3] pioneered the approach in the recreation research, and subsequently many researchers have applied it to the economics of water quality enhancement. Some of these studies are reviewed in Chapter II. Presumably, willingness-to-pay incorporates option demand and aesthetic benefits as well as the benefits from actual recreation.

Bias in benefit estimates from willingness-to-pay surveys are well known, but operate both to over- and under-state the goods' true value. The "free rider" problem suggests that willingness-to-pay will understate the true social value of the good. In the other direction, the fact that the willingness-to-pay debts will never come due could lead to extravagant estimates of value. To our knowledge, no research has adequately sorted out the relative magnitude of these effects.

Three questions were designed to elicit the willingness of respondents to pay for clean water for recreation:

WTP1

A. How much could the cost of visiting this site be raised before you started visiting your second most favorite site more:

a. $.50
b. $1.00
c. $2.00
d. $3.00
e. $4.00
f. $5-10.00
g. more than $10.00

WTP2

B. Suppose that this site were to become very polluted and the water quality would be reduced to a ranking of 1. This could be avoided if sufficient funds were raised to pay for the necessary clean-up. If these funds were to be raised through a higher entrance fee, how much would you be willing to pay to prevent this decline in water quality?

a. $.50
b. $1.00
c. $2.00
d. $3.00
e. $4.00
f. $5-10.00
g. more than $10.00

WTP3

C. Suppose that the water quality could be made much better (improved to a ranking of 5) if sufficient funds were raised to pay for the necessary clean-up. If these funds were to be raised through a higher entrance fee, how much would you be willing to pay to achieve the water quality improvement?

a. $.50
b. $1.00
c. $2.00
d. $3.00
e. $4.00
f. $5-10.00
g. more than $10.00

The analysis of these questions is in four parts. The first section below outlines the principal theoretical underpinning need for interpreting the responses to those questions. The next section analyzes the responses to the three questions via contingency tables. Mean willingness-to-pay is computed, and variations across subgroups of the sample are examined. Contingency tables are too restrictive to examine adequately the determinants of willingness-to-pay on the possible non-linear functional relationships involved. The third section uses OLS regression to probe those relationships more deeply. The final part of this chapter summarizes the major empirical findings and presents some benefit estimations based on these findings.

1. <u>The Theoretic Basis for Willingness-to-Pay Calculations</u>

Three measures of willingness-to-pay are available, corresponding to the three survey questions reproduced above. This brief and informal explanation of the theoretical infrastructure underlying these concepts is intended to define more precisely what these questions measure and the distinctions between them. A more formal analysis of willingness-to-pay (consumer surplus) and specification of the demand curve is presented in Chapter VII below.

The analysis starts with the individual's demand curve for a given site, which we assume to be a function of some measure of the cost of recreation at the site (including travel cost, entry fee, etc.); we use the blanket term "price" to refer to this variable. Temporarily ignoring the other variables which might affect the demand for the site, draw the individual's demand curve as a function of the price of the site; this curve is represented by the line DD' in Figure VI-1a. In this diagram, the recreationist is assumed to face a price of OP for visiting the site and, at that price, he makes OQ visits. Following the standard agrument of elementary micro-economic testbooks, we assert that the area OPDAQ may be taken as an approximate measure of the consumer's total benefit from making OQ visits to the site, the area OPAQ measures his expenditures for visiting the site, and the area PDA may be taken as an approximate measure of his <u>net</u> benefit (consumer's surplus) from visiting the site OQ times. This last area is (approximately) the maximum additional amount which the individual would be willing to pay for visiting the site OQ times rather than not at all.

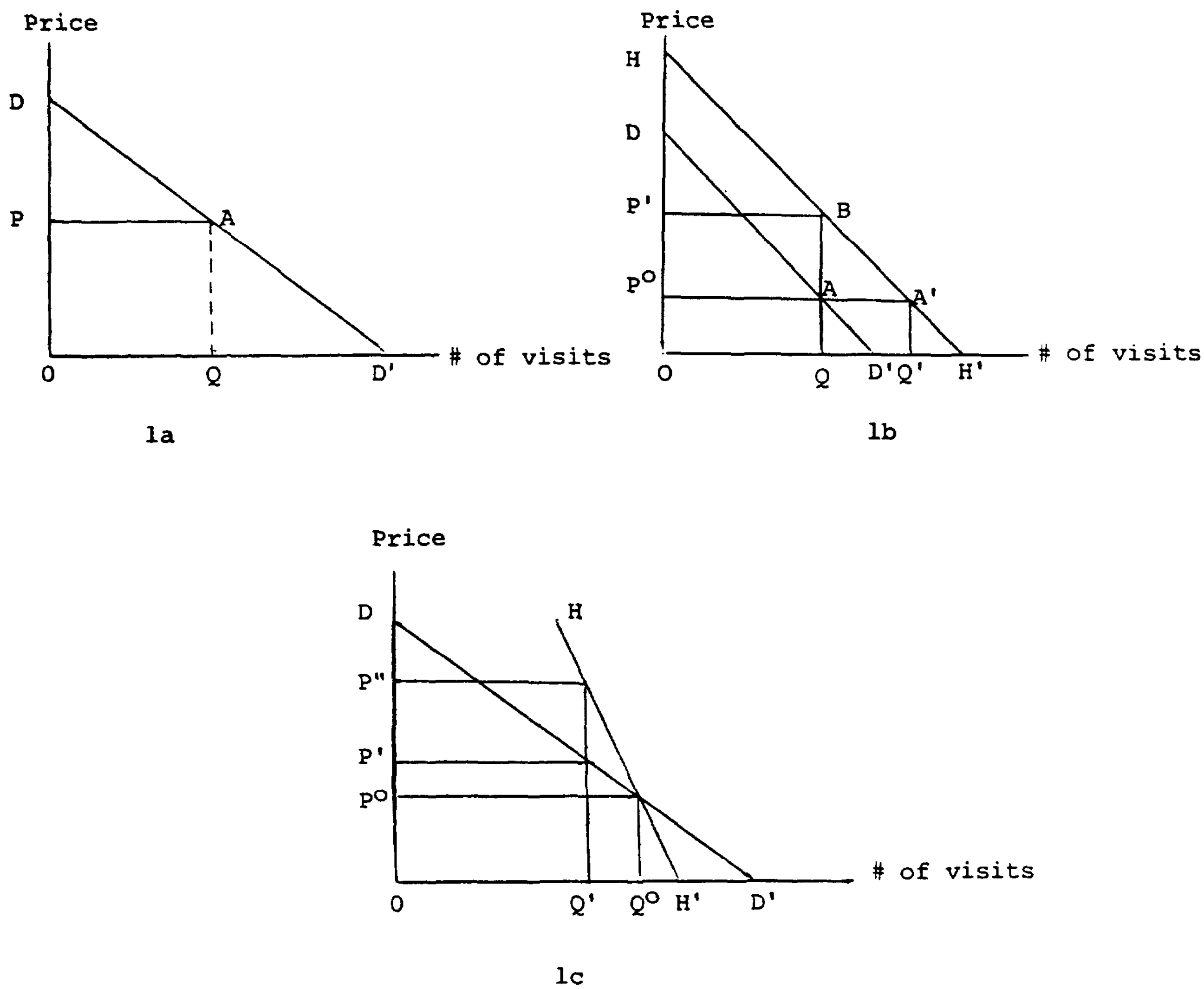

Figure VI-1: Demand Curves for an Individual Recreation Site

What can be said about the determinants of this area? Holding all other variables constant, it is larger when the price of the site is lower (and the number of visits to the site larger). It will also be affected by variables which shift the demand curve DD' holding price constant. Thus, if recreation at the site is a normal good and the individual's income rises, the demand curve would shift outwards. This is illustrated in Figure VI-1b. If the individual's income rises (or if we are comparing two individuals, one having a larger income than the other) the demand curve changes from DD' to HH'; with price constant at OP, the net benefit increases, the amount of the increase being the area ADHBA'. Similarly, if some alternative site which the individual might visit as a substitute declines in quality, we would expect the individual's demand for this site to increase and, with it, net benefit. Finally, if the quality of this site itself is upgraded, we would expect his demand to increase; assuming his demand curve shifts from DD' to HH' we may take the area as an approximate measure of his net benefit from the improvement in quality. Conversely, if the site's quality declines and if the initial demand curve is taken to be HH', this area is a measure or approximate measure of the disbenefit arising from the quality change. Probably it is a function of the magnitude of the quality change, but not necessarily of other variables. However, it is possible that this area is a function of the <u>initial</u> level of water quality or the initial number of visits (if we assume, say, a declining marginal utility of water quality) and it is not inconceivable that it is also a function of income (if we assume that the marginal utility of site quality is not constant with respect to income). Nevertheless it is quite possible that these variables might not affect the magnitude of the net benefit for water quality changes.

With this background, we can consider more precisely what the willingness-to-pay questions measure using Figure VI-1b. Consider the last measures, WTP3, the value of achieving water quality increases is assessed. Here we ask the respondent to tell us the maximum he would pay (i.e., $P^{O}P'$) to move his demand curve from DD' to HH' (and implicity still consume OQ units of recreation). His net benefit before and after the shift must be equal (or else he would be willing to pay more) so the areas P'BH and $P^{O}AD$ must be equal. The net benefit he would receive if water were improved and the charges not levied is, therefore, $P^{O}A'BP'$. This quantity is proportional to $P^{O}P'$ and an estimate which understates its magnitude is given by $P^{O}ABP'$. Of course, this analysis assumes the demand curves are approximately linear over the range considered and that DD' and HH' are parallel. Note that the parallel shift assumption is the more critical one for recovering reasonable approximations to the change in net benefits from the willingness-to-pay questions.

Ideally, we would like to determine the willingness-to-pay over the entire season rather than the willingness-to-pay per visit and then an exact measure of net benefit would be available. But, the former is manifestly unreliable in a survey research context. For any respondent, willingness to pay over the whole season can be estimated by multiplying the reported willingness-to-pay by number of the current visits. WTP2, the value of avoiding water quality declines can be derived similarly.

Figure VI-1c has been constructed to help analyze the first willingness-to-pay question, WTP1. This question asks how much the cost per visit could be increased before the number of visits declines, not necessarily to zero, but to some smaller number, and the best substitute for that site is visited more often. When perfect substitutes are available, consumers' surplus vanishes. This question in effect uses the implicit rates of substitution between the two more preferred sites to compile the net benefit of the most preferred site.

If the consumer is presently visiting the site Q^o times, we assume that if he visits it less he visits it Q' times where Q^o-Q' is some integer (not necessarily unity) which depends on the relative attractiveness of this site and the second most favorite site. The situation is depicted in Figure VI-1c for two different demand curves, DD' and HH'. Suppose, first, that the true demand curve is DD'; with price P^o, the individual makes Q^o visits. The question, in effect, asks for the maximum length $(P'-P^o)$ such that if price increased to P', the individual would begin to reduce the number of his visits.

The change in net benefit from this change in price and consumption equals P^oQ^oBP'. In general, this area depends on the magnitude of the "minimum required reduction" Q^oQ', which is unknown to us. Assume the reduction is small (i.e., Q^oQ' equals unity) which is not implausible given the wording of the question. Then the change in net benefit is bounded above by the quantity $(P^oP')\cdot Q^o$, the reported willingness-to-pay multiplied by number of visits prior to the price increases.

Observe that the net benefit depends strongly on the slope of the demand curve. To see this compare the demand curves DD', and HH' in Figure VI-1c. With the latter demand curve, the same starting amount, and "minimum required reduction," the answer to our question would be P^oP'', a considerably larger amount than P^oP'. But under those conditions, and assuming that the demand curve is linear over this range, then the percent error in the net benefit estimate does not depend on the slope of the demand curve.

We hypothesize that the magnitude of the price increase ($P^{0}P'$ or WTP1) is positively related to the respondents household income and the quality of the site, and negatively to the price of visiting the site (measured by, say, travel time or distance). It may be positively or negatively related to the total number of visits to the site and the total number of visits to other sites.

2. Tabular Analysis of Willingness-to-Pay

The responses to the willingness-to-pay questions are presented in Table VI-1. Several results from this table are of interest. First, the mean values of willingness-to-pay is greater than zero (significant at the 5% level) for all three measures. In other words, despite their inaccurate perception of water quality, respondents were willing to pay to avoid it. This suggests that the principal benefits of water quality improvements are essentially "conservation" oriented rather than "use" oriented.

Second, the incremental value of the favorite site over the second site is less than the value of either avoiding water pollution or achieving water quality improvements (the difference is not, however, statistically significant at the 5% level). Since to avoid the water quality deterioration, the person could shift to the second site and not pay the added cost, this difference reinforces the hypothesized non-usage (merit good, latent demand, option demand, or aesthetic) benefit of water quality improvement. In fact, since we have found only tenuous, at best, support for the relationship between water quality and recreation behavior, we might speculate that most of the willingness-to-pay is in these categories.

The third result is that willingness-to-pay is symmetric between avoiding declines and achieving improvement in water quality. A three-way contingency table shows a strong correlation between response to WTP2 and WPT3 (i.e., the hypothesis of independence can be rejected at the 5% level). This is not unexpected in survey research. Furthermore, the distribution means for WTP2 and WTP3 are nearly identical and the standard deviation differs only by 1.1%, largely because most respondents answered the questions identically. This similarity suggests two hypotheses: either tastes are symmetric and the water quality

TABLE VI-1

Distribution of Willingness to Pay
($ per visit)

	Question					
	WTP1		WTP2		WTP3	
	#	%	#	%	#	%
$ .50	128	36.0	86	25.4	84	24.8
$1.00	84	24.0	113	33.3	113	33.3
$2.00	68	19.4	67	19.8	70	20.6
$3.00	26	7.4	26	7.7	27	8.0
$4.00	17	4.9	14	4.1	14	4.1
$5-10	16	4.6	24	7.1	21	6.2
> 10	13	3.7	9	2.7	10	2.9
Median	1.083		1.239		1.237	
Mean	1.978		2.077		2.034	

rating equals 2.5 or tastes are nonsymmetric to account for water quality ratings different from 2.5. As seen in Section IV 2.3 above, the mean water quality rating equals 2.881, and is slightly skewed to the right. A rating of 2.5 is not statistically different (at 5% confidence) from the observed mean. Combined with the symmetry of response to questions WTP2 and WTP3, the difference suggests avoiding water quality declines is not so valuable as achieving water quality improvements. This is contrary to the expressed preferences which associated (negatively) site choice only with bad water quality and find little if any response to good water quality. Again, we must conclude that these willingness-to-pay questions measure something outside recreational usage.

Previous studies have found willingness-to-pay for water quality improvement to be related to income and education. Our analysis is more limited, being confined to the recreation context, but we still would expect a positive correlation between willingness-to-pay and income, education and occupation. Too, we expected whites to have higher levels of willingness-to-pay than blacks. None of these hypotheses were confirmed at the 5% level.* No S-shaped curve between income and willingness-to-pay, as suggested by some authors could be discerned from the tables. A significant positive correlation was found between family size and willingness-to-pay, but this relationship disappeared when willingness-to-pay was computed on a per capita basis.

This absence of correlation was surprising. Since our sample SES characteristics are close to those for the SMSA as a whole, these results suggest that the willingness-to-pay is uniform across the population. The individual amounts are small, so perhaps they do not constitute an adequately large portion of total income to induce any differential effect.

*The next section probes these relationships in greater depth.

Alternatively, in general, the poorer group of our sample live closer to the lower quality inner city beaches. Conversely, the more wealthy visit the better quality outer beaches more often. Since there was substantial agreement concerning the perceived water quality across the sites, we could postulate that the poor are willing to pay more in proportion to their income than the wealthy because they currently visit poorer sites and would like to see them improve. However, then the wealthy should be willing to pay more to avoid declines in their good sites and a positive income correlation with WTP2 should exist. But no such correlation was found. Bolstered by the regression analysis in Section 3, Section 4 of this chapter returns to these conclusions.

A second set of hypotheses were formulated to examine the relationship between willingness-to-pay and access to recreation. Access included ownership of an automobile, amount of leisure time each week, amount of vacation time per year, total amount of recreation equipment owned and the use of public transit. We expected auto ownership to be negatively correlated and all the others positively with willingness-to-pay. At the 5% level, only transit usage was significant as shown in Table VI-2. Frequent users of public transit may not have access to high quality sites, and, therefore, perceive greater benefits from water quality improvements and disbenefits from declines.

The last subgroup examined were participants in various activities. We hypothesized that participants would be more sensitive to water quality benefits than non-participants. For swimmers, boaters, walkers and bicyclists, the hypothesis was not proved. For fishermen, the hypothesis can be accepted at a 5% level of confidence, and the contingency table is shown in Table VI-3.

Table VI-2

Willingness to Pay By Transit Usage

	Transit Use			
	Never	Almost Never	Occasionally	Frequently
a. $.50	(12) 14.1 18.2	(14) 16.5 19.2	(7) 8.2 12.1	(52) 61.2 36.9
b. $1.00	(23) 20.4 34.8	(27) 23.9 37.0	(27) 23.9 46.6	(36) 31.9 25.5
c. $2.00	(11) 16.4 16.7	(20) 29.9 27.4	(11) 16.4 19.0	(25) 37.3 17.7
d. $3.00	(7) 26.9 10.6	(5) 19.2 6.8	(6) 23.1 10.3	(8) 30.8 5.7
e. $4.00	(1) 7.1 1.5	(2) 14.3 2.7	(3) 21.4 5.2	(8) 57.1 5.7
f. $5-10.00	(11) 45.8 16.7	(3) 12.5 4.1	(2) 8.3 3.4	(8) 33.3 5.7
g. more than $10.00	(1) 11.1 1.5	(2) 22.2 2.7	(2) 22.2 3.4	(4) 44.4 2.8

Table shows cell count in ( ), row percentages and column percentages.

Table VI-3

Willings to Pay by Participation in Fishing

	1	2	3	4	5	6	7
1. Fishermen	(40) 28.2 31.7	(39) 27.5 46.4	(26) 18.3 38.2	(16) 11.3 61.5	(9) 6.3 52.9	(5) 3.5 31.3	(7) 4.9 53.8
2. Non-Fishermen	(86) 41.3 68.3	(45) 21.6 53.6	(42) 20.2 61.8	(10) 4.8 38.5	(8) 3.8 47.1	(11) 5.3 68.8	(6) 2.9 46.2

Table shows cell count in ( ), row percentages and column percentages.

3. Regression Analysis of Willingness to Pay

For ordinary least squares regression analysis, it is convenient to continuous variables for both the dependent variable--willingness to pay--and the independent variables. This assumption is not strictly necessary--we shall relax it partially below--but it greatly simplifies the analysis and it seems to be fairly reasonable in the present case. The answers to the willingness to pay questions are essentially ranges: the respondent who checks response (d)--$3--may be presumed to be actually willing to pay some amount greater than $2.50, but less than $3.50, and similarly with the other responses. Nevertheless, the ranges are relatively small, and therefore it is not unreasonable to use the midpoints of the ranges in place of the unknown means. A similar argument applies to the income variable. In doing this we arbitrarily take the (unknown) midpoint of the last willingness to pay answer--"more than $10--to be $15 and with the income variable we take the midpoint of the first income class to be $2,500 and that of the last class to be $60,000.*

The properties of the resulting estimator have been analyzed by Haitovsky [4]. He shows that they are biased in general, but if the number of categories into which the dependent variable is classified is the same as the number of categories into which the explanatory variable is classified, the resulting estimator will be the same as that obtained by using the (unknown) <u>means</u> of the ranges instead of the midpoints. Cramer [2] has shown that the latter estimator is unbiased, although inefficient. Haitovsky [4] also shows that when the number of categories for the explanatory variable is larger than for the number for the dependent variable--

*These values are actually closer to the mean of the first and last groups computed from a Pareto distribution.

as is the case when we regress willingness to pay on income--the slope coefficient obtained by using the midpoints is likely to be larger in absolute value than that obtained by using the means. In addition, he shows that the loss of efficiency due to grouping declines as the category size is smaller and as the population correlation between the dependent and independent variable approaches unity.

The other issue which we must address is the functional form of the relationship between willingness to pay and its determinant. We had no reason a priori to prefer any particular form. We therefore considered several different functional forms, including the following:

I	$\frac{1}{y} = a + b/x.$	$\varepsilon_{yx} = \frac{b}{(ax + b)}$
II	$\ln y = a - b/x$	$\varepsilon_{yx} = b/x$
III	$\ln y = a + bx$	$\varepsilon_{yx} = bx$
IV	$\ln y = a + b \ln x$	$\varepsilon_{yx} = b$
V	$y = a + b \ln x$	$\varepsilon_{yx} = \frac{b}{y}$
VI	$y = a + bx$	$\varepsilon_{yx} = \frac{bx}{y}$

where $\varepsilon_{yx} = \frac{dy}{dx}\left(\frac{x}{y}\right)$ is the elasticity of y with respect to x.

Form I, with $b<0$, is an shaped function, intercepting the x-axis at zero and approaching $1/a$ asymptotically as x increases to infinity. Form II with $b>0$ is an shaped function, passing through the origin and approaching (a) asymptotically as x increases to infinity. Form III with $b>0$ is an shaped function, cutting the y-axis at a. Form V with $b>0$ is shaped rather like Form I, except that it cuts the x-axis at $e^{-a/b}$ and increases without bound as x increases. The shapes of the other two functions require no explanation. When necessary, an appropriate criterion for choosing among alternatives II, III and IV, or between V and VI is minimizing the residual sum of squares from the fitted regression--or, equivalently, maximizing the R^2 statistic. However, in order to choose between the three broad classes of functions (I), (II, III, IV), (V, VI), with respectively 1/y, 1n y andy as the dependent variable, it is necessary to apply the likelihood ratio test suggested by Box and Cox [1].

As before, we refer to the additional willingness to pay for visiting the respondent's favorite site as WTP1, the willingness to pay to prevent the site from becoming polluted as WTP2, and the willingness to pay to obtain a higher level of water quality as WTP3. Since these three measures pertain to different concepts, there is no reason why they should be identical in value. In order to test this, we regress one measure on the other; if the two measures were identical, the estimated intercept would not be significantly non-zero and the estimated slope coefficient would not be statistically different from unity. The regressions are performed on the data subsets containing answers to both questions, for each of the three pairs of measures. The results are as follows:[1]

$$\underset{}{WTP2} = \underset{(5.88)}{1.031} + \underset{(11.73)}{0.5715}\ WTP1 \qquad R^2 = .335 \qquad (275\ \text{obs.})$$

$$\underset{}{WTP3} = \underset{(5.94)}{0.983} + \underset{(12.48)}{0.547}\ WTP1 \qquad R^2 = .362 \qquad (277\ \text{obs.})$$

$$\underset{}{WTP3} = \underset{(2.55)}{0.248} + \underset{(31.4)}{0.8662}\ WTP2 \qquad R^2 = .772 \qquad (293\ \text{obs.})$$

[1]The numbers in parentheses below the coefficient estimates are t-statistics.

Clearly, WTP2 and WTP3 are closer in value to each other than to WTP1, but no pair of these measures is sufficiently close to be considered statistically identical.[1]

Determinants of WTP1

On the basis of the considerations outlined in Section 1, we hypothesize that WTP1 is a positive function of income (INC), a negative function of travel time (TIME) and distance to the site (DIST), which are a large component of the site's "price", a positive function of the household's total number of visits to the site (HVS), and a positive function of the site's quality. For the last variable we can use either the respondent's subjective rating of the site's characteristics or the "objective" water quality characteristics.

The results of some bivariate regressions are shown in Table VI-4. It turns out that there is little relationship between willingness to pay and income. The two preferred equations--one of them representing an S-shaped relationship--indicate that the relationship is significant at the 90%, but not the 95% level. As hypothesized, there is a positive relationship between the number of visits and willingness to pay. Willingness-to-pay and travel time or distance, which may be taken as proxies for price, are also positively associated, an unexpected result. We discuss this result in greater detail below.

The next three sets of regressions show that there is a strongly significant relationship between willingness-to-pay and

---

*In fact, our of the 293 cases where the respondent provided data on both WTP2 and WTP3, the response was the same in 251 cases: in 24 cases WTP3 exceeded WTP2 and in 18 cases WTP2 exceeded WTP3.

All the intercepts are significantly different from zero, and slope coefficients are less than unity at the 95% level.

Table VI-4

Some Regressions with WPT1 as Dependent Variable

	Form	Equation	R^2	f
	INCOME (256 observations)			
*	FORM I.	1/WTP1 = 0.9854 + 988.47/INC (13.37) (1.78)	R^2=.012	f=3.12
*	FORM II.	1n(WTP1) = 0.3672 - 1177.28/INC (1.64)	R^2=.01	f=2.68
	FORM IV.	1n(WTP1) = -1.017 + 0.13431n(INC) (1.19) (1.49)	R^2=.009	f=2.21
	HOUSEHOLD VISITS TO SITE (308 observations)			
	FORM V.	WTP1 = 1.57 + 0.31141n(HVS) (4.63) (2.01)	R^2=0.13	f=4.05
*	FORM VI.	WTP1 = 1.92 + 0.0191 · HVS (9.33) (2.05)	R^2=.013	f=4.19
	DISTANCE FROM SITE (290 observations)			
	FORM I.	1/WTP1 - 1.02 + 0.1535/DIST (19.12) (1.741	R^2=.01	f-3.02
*	FORM VI.	1n(WTP1) = 0.029 + 0.0274 · DIST (0.38) (3.87).	R^2=.049	f=14.95
	FORM VI.	1n(WTP1) = 1.767 + 0.0442 · DIST (7.07) (1.92)	R^2=.013	f=3.67
	TRAVEL TIME (293 observations)			
	FORM I.	1/WTP1 = 0.979 + 1.395/TIME (16.7) (1.79)	R^2=.011	f=3.19
*	FORM IV.	1n(WTP1) = -0.405 + 0.20661n(TIME) (1.93) (3.38)	R^2=.038	f=11.43
	FORM VI.	WTP1 = 1.82 + 0.00995TIME (7.12) (2.13)	R^2=.015	f=4.52
	RATING OF WATER QUALITY (303 observations)			
	FORM I.	1/WTP1 = 0.8441 + 0.519/RWQUAL (11.52) (3.50)	R^2=0.39	f=12.27
*	FORM III.	1n(WTP1) = -0.2105 + 0.1485 · RWQUAL (1.59) (3.97)	R^2=.05	f=15.78
	FORM VI.	WTP1 = 1.141 + 0.3223RWQUAL (2.63) (2.63)	R^2=.022	f=6.91
	RATING OF BEACH QUALITY (303 observations)			
	FORM I.	1/WTP1 = 0.9085 + 0.4698/RBQUAL (11.85) (2.48)	R^2=.02	f=6.13
*	FORM II.	1n(WTP1) = -0.295 + 0.1535RBQUAL (1.85) (3.67)	R^2=.043	f=13.46
	FORM IV.	WTP1 = 0.761 + 0.3881RBQUAL (1.46) (2.85)	R^2=.026	f=8.1

Table VI-4 (CONTINUED)

Some Regressions with WPT1 as Dependent Variable

RATING OF CROWDING (308 observations)

FORM I.	1/WTP1 = 0.965 + 0.2273/RCROWD (12.67) (1.59)	R^2=.008	f=2.51
* FORM III.	1n/WTP1) = -0.0197 + 0.094RCROWD (0.15) (2.41)	R^2=.019	f=5.82
FORM VI.	WTP1 = 1.296 + 0.2906RCROWD (2.31)	R^2=.017	f=5.35

FACTOR 4 (245 observations)

FORM I.	1/WTP1 = 1.0277 - 0.00924/FACT4 (21.48) (1.98)	R^2=.016	f=3.91
* FORM II.	1n(WTP1) = 0.2943 + 0.0129/FACT4 (4.9) (2.2)	R^2=.019	f=4.83
FORM VI.	WTP1 = 2.131 + 0.5088 · FACT4 (2.31)	R^2=.021	f=5.33

pH (245 observations)

FORM I.	1/WTP1 = 1.033 + 0.0125/pH (20.92) (1.4)	R^2=.008	f=1.96
* FORM II.	1n(WTP1) = 0.2902 - 0.0186/pH (4.68) (1.66)	R^2=.011	f=2.76

TURBIDITY (187 observations)

* FORM I.	1/WTP1 = 0.965 + 0.5268/TURB (12.24) (3.4)	R^2=.059	f-11.57
FORM IV.	1n(WTP1) = -0.24 + 0.28221n(TURB) (3.15) (3.79)	R^2=0.72	f=14.38
FORM VI.	WTP1 = 1.101 + 0.1468 · TURB (3.79) (3.2)	R^2=.052	f=10.23

COLIFORM BACTERIA (245 observations)

* FORM III.	1n(WTP1) = 0.2036 + 0.0000341 · CBACT (3.33) (1.84)	R^2=.014	f=3.37
FORM VI.	WTP1 = 1.8802 + 0.0000135 · CBACT (9.99) (2.36)	R^2=.022	f=5.56

NOTES:
1. The absolute values of the t-statistic are given in parentheses below the coefficient estimates.
2. The critical values at the 95% level for the t- and f-statistics are respectively 1.96 and 3.84.
3. An asterisk denotes the functional form which is preferred on the basis of the likelihood ratio test.

perceived site quality, as measured by the rating of water quality, beach quality and crowding.* However, the relationship between willingness to pay and "objective" water quality is tenuous at best. Many objective water quality measures, such as the sites' scores for Factors 1, 2 and 3 and such variables as alkalinity and color bear no significant relationship to willingness to pay. Those variables which do have a significant slope coefficient, such as pH (measured in terms of squared deviations from the value of 7), turbidity and coliform bacteria, have a positive coefficient instead of a negative one (it should be remembered that larger values of these variables signify a greater degree of pollution). The only exception is the site scores for Factor 4 (which are positively correlated with bacteria counts); the regressions equation using Forms I and II indicate a significant negative relationship with willingness to pay, while the equation using Form VI indicates a significant positive relationship. This last result is difficult to interpret since it is unlikely that recreationists can perceive bacteria, let alone a composite water quality factor which loads heaving on the bacteria count.

The divergence between the results obtained using subjective ratings of site characteristics and objective measures of water quality reaffirm one's doubts concerning the accuracy of the respondent's perception of water quality conditions at the Boston area sites.

There remains the question of the positive slope coefficient in the regressions of WTP1 on TIME and DIST. Larger values of these variables, signifying a higher cost of access to the site, and should be associated with smaller amounts of willingness to pay. One explanation for the positive slope coefficients is that the more distant sites are of a better quality than the closer sites, so that distance is serving as a proxy for site quality. That this

---

*These variables are here treated as being continuous, cardinal variables, The appropriateness of this assumption was discussed more fully in Section V 2.3, above.

explanation has some validity is shown by the correlation coefficients between distance and various site quality variables displayed in Table VI-5. * In order to examine the relationship between willingness to pay and distance, allowing for the separate effects of site quality, consider these regressions of WTP1 on both distance and quality variables:**

$$\ln(WTP1) = \underset{(1.9)}{-0.323} + \underset{(.69)}{0.0301}\ DIST + \underset{(3.89)}{0.1617}\ RWQUAL$$
$$R^2 = .066 \qquad F = 9.03$$

$$\ln(WTP1) = \underset{(1.51)}{-0.288} + \underset{(.73)}{0.0329}\ DIST + \underset{(2.73)}{0.1328}\ RBQUAL$$
$$R^2 = .039 \qquad F = 5.18$$

$$\ln(WTP1) = \underset{(2.07)}{-0.296} + \underset{(1.59)}{0.0024}\ TIME + \underset{(3.62)}{0.1509}\ RWQUAL$$
$$R^2 = .073 \qquad F = 10.123$$

$$\ln(WTP1) = \underset{(1.64)}{-0.289} + \underset{(2.07)}{0.0031}\ TIME + \underset{(2.69)}{0.1253}\ RBQUAL$$
$$R^2 = .053 \qquad F = 7.13$$

It seems from these regression equations that, even when the effects of site quality are removed, there is still a somewhat positive relationship between willingness to pay and distance. The same conclusion holds when income, which is positively correlated with both distance and willingness to pay, is held constant, as can be seen from the following regressions:***

---

*These correlation coefficients are computed from the full set of data on household visits to all sites, rather than merely the visits to the favorite site.

**These regressions are based on 260 observations; the notation and display is the same as in Table VI-1.

***These regressions are based on 226 observations.

Table VI-5

Correlation of Time and Distance Travelled to 29 Sites With Site Quality Variables

	TIME	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	Bacteria	Color	Turbidity	Rating of Water Quality	Rating of Beach Quality	Rating of Crowding	Household Income
Distance	.453	.218	-.093	-.214	-.210	-.304	-.347	-.263	.395	.202	.278	.125
In (Distance)	.407	.208	-.129	-.232	-.235	-.347	-.376	-.275	.371	.185	.252	.155
Time	1.000	.087	-.099	-.130	-.109	-.179	-.148	-.135	.110	.072	.119	.053

NOTE: "Bacteria" is the arithmetic mean of Coliform and Total Bacteria.

All the correlation coefficients are significantly different from zero at the .01 level; there are about 900 degrees of freedom for all but the last column, for which there are about 750 degrees of freedom.

$$\ln(WTP1) = \underset{(1.31)}{-0.242} + \underset{(1.56)}{0.0139}\ DIST + \underset{(2.69)}{0.1301}\ RWQUAL - \underset{(.48)}{353.15}/INC$$

$$R^2 = .076 \qquad F = 6.1$$

$$\ln(WTP1) = \underset{(1.31)}{-0.242} + \underset{(1.56)}{0.0024}\ TIME + \underset{(3.12)}{0.1422}\ RWQUAL - \underset{(.71)}{520.18}/INC$$

$$R^2 = .076 \qquad F = 6.1$$

Thus, it seems possible that respondents place a positive premium on more distant sites, even when the effects of site quality and income are removed. There are two possible explanations for this phenomenon. The most obvious explanation is that respondents visit those sites for their natural setting, lack of crowding, or other site characteristics not included. Another explanation is based on the specialized definition of the WTP1 variable, discussed in Section 1 above; it may be that the length (Q^0-Q^1) in Figure VI-1c is larger for more distant sites than for nearer sites; that is, if the household is to reduce the number of its visits to its favorite site, the minimum reduction is larger for more distant sites. The alternative explanation is that recreation sites, like certain other commodities,. may be subject to the Veblen effect: consumers are willing to buy larger quantities of the higher priced good.

Determinants of WTP2

The results of some regressions of WTP2 on various explanatory variables are shown in Table VI-6. Willingness to pay to avoid very polluted site condition appears to be an increasing function of income, although the confidence intervals on this result are wide. Also, increases in present site conditions tend to increase WTP2. From the fact that functional form III has the best fit of all six forms, we may infer that willingness to pay elasticity actually increases with quality of present site conditions, which refutes the diminishing marginal utility of water quality hypothesis

Table VI-6

Some Regressions with WTP2 as Dependent Variable

INCOME (247 observations)

I.	1/WTP2 = 0.8492 + 1253.57 · 1/INC	R^2=.033	f=5.42
	(12.26) (2.33)		
III.	ln(WTP2) = 0.185 + .00001113 · INC	R^2=.013	f=3.3
	(1.7) (1.82)		
IV.	ln(WTP2) = -1.1571 + .1602 ln(INC)	R^2=.013	f=3.2
	(1.37) (1.79)		

RATING OF WATER QUALITY (292 observations)

I.	1/WTP2 = 0.825 + 0.3828/RWQUAL	R^2=.025	f=7.32
	(11.68) (2.7)		
III.	ln(WTP2) = -0.0549 + .1188 · RWQUAL	R^2=.034	f=10.32
IV.	ln(WTP2) = 0.05083 + 0.2658ln(RWQUAL)	R^2=.027	f=8.05
	(.46) (2.84)		

RATING OF BEACH QUALITY (294 observations)

I.	1/WTP2 = 0.8109 + 0.529/RBQUAL	R^2=.032	f=9.96
	(11.39) (3.11)		
III.	ln(WTP2) = -0.162 + 0.1319 · RBQUAL	R^2=.037	f=11.1
	(1.08) (3.33)		
IV.	ln(WTP2) = -0.0338 + 0.29ln(RBQUAL)	R^2=.025	f=7.36
	(.25) (2.71)		

PARTICIPATION IN FISHING/BOATING (303 observations)

III.	ln(WTP2) = 0.209 + 0.2094PART	R^2=.014	f=4.26
	(2.80) (2.06)		
VI.	WTP2 = 1.838 + 0.6497PART	R^2=.014	f=4.16
	(7.85) (2.04)		

DISTANCE/TIME, WATER QUALITY RATING AND INCOME (226 observations)

III.	ln(WTP2) = -1.113 + 0.0116DIST + 0.0815RWQUAL + 0.117ln(INC)	R^2=.05	f=3.92
	(1.29) (1.29) (1.72) (1.25)		
III.	ln(WTP2) = -1.276 + 0.000794TIME + 0.1014 RWQUAL + 0.1344ln(INC)	R^2=.045	f=3.46
	(1.49) (0.51) (2.28) (1.46)		

suggested above. In addition, we have regressed WTP2 on a dummy variable PART, which takes the value 1 of members of the respondents' household engaged in boating and/or fishing, and the value 0 otherwise. As we might expect, participation in these activities increases the respondent's willingness to pay to avoid pollution by about 20% over nonparticipants. Finally, as with WTP1, there is some evidence of a positive relationship between distance and willingness to pay, even when water quality rating and income are held constant.

Determinants of WTP3

The results of some regressions of WTP3 on several explanatory variables are shown in Table VI-7. The most important finding is that willingness to pay to obtain an improvement in water quality increases with present site quality. This is completely counterintuitive: we had hypothesized that willingness to pay would be greatest when existing site conditions were very poor, because visitors to such sites would have the greatest amount to gain, both absolutely and relative to the starting position. The finding that the reverse seems to be true suggests that the taste for water quality increases with the respondent's exposure to it. In terms of utility theory, we are suggesting that the marginal utility of water quality may increase with "consumption" of water quality, at least within the range covered by the present sample.

Table VI-7

Regressions with WTP3 as Dependent Variable

INCOME (247 observations)

I.	$1/WTP3 = 0.9007 + 736.33/INC$ (12.97)	$R^2=.008$	F=1.87
III.	$ln(WTP3) = 0.2878 + 0.00000344\ INC$ (2.71) (.57)	$R^2=.001$	F=0.33
IV.	$ln(WTP3) = -0.1737 + 0.0544 ln(INC)$ (.21) (.62)	$R^2=.002$	F=0.38

RATING OF WATER QUALITY (292 observations)

I.	$1/WTP3 = 0.819 + 0.3544/RWQUAL$ (11.77) (2.61)	$R^2=.023$	F=6.81
III.	$ln(WTP3) = 0.0093 + 0.1023\ RWQUAL$ (.07) (2.82)	$R^2=.027$	F=7.97
IV.	$ln(WTP3) = 0.1058 + 0.2229 ln(RWQUAL)$ (1.0) (2.46)	$R^2=.02$	F=6.03

RATING OF BEACH QUALITY (295 observations)

I.	$1/WTP3= 0.7908 + 0.5281/RBQUAL$ (11.38) (3.24)	$R^2=.035$	F=10.5
III.	$ln(WTP3) = -0.088 + 0.116\ RBQUAL$ (.60) (2.98)	$R^2=.029$	F=8.87
IV.	$ln(WTP3) = 0.0171 + 0.2616 ln(RBQUAL)$ (.13) (2.51)	$R^2=.021$	F=6.31

4. Conclusions: Dollar Values of Willingness to Pay in the Boston SMSA

Willingness to pay for water quality exceeds zero despite the generally poor perception of water quality. The evidence suggests that the net benefits implied by this do not necessarily derive from the direct usage of the water, but may also be based on an option demand character of water quality. Bostonians appear to value conservation.

Willingness to pay to either achieve water quality improvements or avoid water quality degradation increases with better site quality. In other words, the value of improving/maintaining good sites is greater than that for poorer sites. This finding holds once income and distance (setting) effects are removed as well. It suggests there are increasing returns to water quality improvements. Because the costs of water pollution abatement typically display increasing marginal costs, this finding implies that much higher levels of water quality contact than previously thought may be socially efficient.

From the response to the willingness-to-pay questions (WTP2 or WTP3), a dollar value of water quality improvements (or cost of declines) can be estimated from the formula developed in Section 1. Recall these estimates probably overstate the true net benefits. We assume our sample is representative of the Boston SMSA population, and no adjustments are needed to account for variation due to social, economic or other factors. On the average, responding households made 20.75 visits to a recreation site during the period. Valued at the median willingness-to-pay figure (1.259) this implies a value of about 26.11 per household per year for water quality improvements. This equals $17.3 million per year for the 1970 Boston SMSA population. Using the mean figure of $2.065, the per capita figure becomes $42.85 per year, and the SMSA figure rises to 28.4 million per year. Because the data are categorical, confidence bands for these estimates cannot be simply calculated. But the distribution is skewed to the right, so any equal probability confidence intervals would find deviations to the high side more likely. Remember that this value is not necessarily generated by direct recreation usage alone, but also by the conservation value of achieving and maintaining good quality water in the Boston area.

CITED REFERENCES

1. G.E.P. Box and D.R. Cox, "An Analysis of Transformations," Journal of Royal Statistical Association Series B (1964): 211-252.

2. J.S. Cramer, "Efficient Grouping, Regression and Correlation in Engle Curve Analysis., Journal of American Statistical Association (1964): 233-250.

3. R.K. Davis, The Recreation Value of Northern Maine Woods, unpublished Ph.D. thesis, Department of Economics, Harvard University, 1963.

4. Yeol Haitovsky, Regression Estimation from Grouped Observations, 1973.

VII. MULTIPLE SITE DEMAND FUNCTIONS

The formal economic analogue to willingness-to-pay is consumer's surplus measured from an appropriately specified demand function. Our analysis focuses on multiple site demand systems because substitutions between the sites were significant. Table VII-1 shows the response to a direct question on substitutions:

> Let's talk about the beach, lake or river site you visited most, that was ________________, site number ________.
>
> If water quality became much worse (declined to a ranking of 1), what would your response be?
>
> a. still visit the same beach as much
>
> b. visit that site less frequently and some other site more (specify which one below)
>
> c. visit that site less frequently and participate in some other non-water-based recreation more (specify which activity below).
>
> d. participate in outdoor recreation less, no change in other leisure
>
> e. participate in outdoor recreation less and indoor recreation more.

Most (56.9%) respondents would shift to their second most favorite site. Over three-quarters of all respondents would continue to participate in water-based activities at the system of sites under study.

Table VII-1

Substitution Induced by Water Quality Decline

	Response	No.	Percent
a.	still visit the same beach as much	83	20.9
b.	visit that site less frequently and some other site more	226	56.9
c.	visit that site less frequently and participate in some other non-water-based recreation more	53	13.4
d.	participate in outdoor recreation less, no change in other leisure	21	5.3
e.	participate in outdoor recreation less and indoor recreation more	14	3.5

Five sections complete the demand analysis. The first section discusses in a qualitative way demand at the system of sites. Section 2 presents some aggregated regressions which focus more specifically on the determinants of recreation behavior. These sections, combined with the background matter presented in previous chapters, set the stage for the demand modelling of sections 3 and 4. Section 3 employs the abstract site demand functions pioneered in transportation economics to estimate the functional relationships between site characteristics and site demand. However, the specification does not permit recovery of an exact measure of consumers' surplus (net benefit), so Section 4 considers a system of demand equations derived explicitly from a utility model. Unfortunately, estimation of these equations, a complex operation, exceeded the level of the project's resources. This model is left specified but not estimated. The last section presents benefit estimates from the abstract site model, and comments on benefit estimates from the system demand model.

1. A Review of the Data

Table VII-2 shows the number of mentions and visits for each site in our survey. The first column contains the number of households who visited each site at least once during the summer of 1974; the second column gives the total number of visits to the site by these households. The median number of visits to a site, computed from the third column of the table, is 7 visits per household. For reasons to be explained below, the statistical analysis will be focused mainly on sites 1-29; these sites account for almost 80% of the total number of mentions but only 66.6% of the total number of household visits. Thus the excluded sites appear to have a somwhat higher average visitation rate per household. In fact, however, this is misleading because some of the excluded sites are really composites of individual sites. If we adjust for this, the average visitation rates for the included and excluded sites would be fairly similar.

To get some feel for the coverage of the sample Table VII-3 presents a comparison of the site attendances generated by the respondents to our questionnaire and estimates of total attendance at selected sites for which data is available. The data in second column of the table was obtained by multiplying the number of household visits to each site by the average group size and summing this over all respondents. The data in the first column comes from a variety of sources. Attendance figures were generally not available at the head office of the MDC or at other official agencies in Boston, but some data was available from staff at the sites when we visited them. The quality of the data is unknown: some of it comes from a survey conducted in 1969; in other cases the data is based on parking and entrance fee receipts. Taking this data at face value, observe that the households in our sample generated 0.13% of the estimated total attendance at these sites. This may be compared with the ratio between our sample population and the total Boston area population, which

Table VII-2

Individual Site Visits and Mentions

Site	# of Mentions	# of Household Visits	(2)/(1)
1	21	150	7.1
2	45	306	6.8
3	98	681	6.9
4	112	906	8.1
5	9	98	10.9
6	15	68	4.5
7	14	188	13.4
8	30	107	3.6
9	7	22	3.1
10	11	44	4.0
11	9	209	23.2
12	11	121	11.0
13	11	57	5.2
14	4	16	4.0
15	30	382	12.7
16	115	948	8.2
17	51	256	5.0
18	74	306	4.1
19	43	167	3.9
20	23	77	3.3
21	15	80	5.3
22	34	212	6.2
23	14	162	11.6
24	17	216	12.7
25	20	312	15.6
26	47	180	3.8
27	8	26	3.3
28	22	143	6.5
29	2	15	7.5
30	24	294	12.3
31	43	556	12.9
*32	12	102	5.7
33	10	71	7.1
*34	4	8	2.0
35	9	94	10.4
36	4	74	18.5
37	4	96	24.0
*38	27	408	15.1
*39	24	300	12.5
*40	18	119	6.6
41	11	141	12.8
*42	49	937	19.1
*43	6	30	5.0
All Sites	1163	1685	8.3
'Mean	2.49	20.74	9.0

NOTE: Column (1) excludes those respondents who mentioned a site for the purpose of rating its characteristics but did not actually visit it.

An Asterisk denotes those "sites" which are actually groups of individual sites; each mention refers to a different individual site and/or different respondent.

Table VII-3

Total Attendance and Attendance from Sample Households At Selected Sites

Site	(1) Estimated Annual Attendance (10^3 visitor days)	(2) Attendance by Sample Households (visitor days)	(3) Percent of Total Attendance Generated by Sample
1		428	
2	2000	957	0.17
3		1998	
4	6400	5124	0.08
5	350	2021	0.58
6	750	289	0.04
7	500	881	0.18
9		92	
10	2500	90	0.02
12		384	
15	750	1628	0.22
16	2700	3370	0.12
18	140	1246	0.89
22	150	918	0.61
23	175	991	0.57
24	750	602	0.08
27	40	84	0.21
28	120	662	0.55
29	105	141	0.13
TOTAL	17,430	21,911	0.13

NOTE: Column (2) is number of visits by household members to sites multiplied by average group size.

Column (3) contains fractions of one percent.

amounts to about 0.06%. The comparison suggests that the households in our sample could be responsible for more recreation visits than the average household in the Boston area. However, this conclusion must be treated with considerable caution, for the total attendance estimates are not reliable. Some of these figures date back to 1969 and others are only guesses of numbers of automobiles, so that they understate present attendance levels. On the other hand, it should be noted that the attendance may have been generated by a population larger than that of the Boston metropolitan area, since they may contain visits by tourists from elsewhere in the state or from out of state.

The next issue to be considered is how many sites each household visits. We pointed out in Chapter III that certain statistical site demand models could be applied only if it were believed that each individual visited one and only one of the alternative sites. It is therefore important to check the validity of this assumption. Table VII-4 shows the distribution of the number of sites visited by respondents. It is clear that the assumption is not valid: two thirds of the sample visited more than one site in the summer of 1974. In fact, that mean number of sites visited was 2.5 sites per household, and the median and modal number was 2 sites. Thus we must rule out those models which presuppose the choice of a single site.

In fact, two types of demand models were estimated. The explanatory variables in one type include income and household structure and the own price and quality variables for the site: in the other type of models, besides these variables, there are also the prices and quantities of the other (n-1) sites. In order to generate the data on subjective site quality ratings necessary for the implementation of the second type of model we included questions in our questionnaire asking respondents to rate the quality of other sites which they knew about but did not visit. Unfortunately, these questions were not very successful and,

Table VII-4

Household Site Visitation Patterns

# of Sites Visited	# of Occurrences
0	56
1	106
2	114
3	69
4	54
5	21
6	17
7	10
8	8
9	3
10	2
11	1

for one reason or another, most respondents did not answer them. Thus, while we have 1312 site cards, each representing the metnion of one site by one respondent, only 148 cards represent the mention of sites which the respondent did not visit but where he was willing to rate site quality. To all intents and purposes, then, we do not have subjective ratings of the sites which respondent did not visit. Since most respondents visited only 2 or 3 sites, this rules out the majority of the sites where we wish to model demand. Accordingly, if we wish to include a full set of (n-1) other site variables in each demand equation, we have to use the objective measure of water quality obtained from our water samples from 29 sites. This is why we are forced to exclude sites 30-43 from most of the statistical anlaysis.

The same problem arises with the price variable. However, there are some additional considerations. The questionnnaire asks how much it costs respondents to gain access to a site in parking or entrance fees. It also asks how much respondents spend once they are at the site. As Table VII-5 shows, most persons said that they incurred no expenditures for access--about 73% of the mentions indicate a zero price--and about one third of the respondents said they had no on-site expenditures. We cannot tell how accurate these responses are: since the interviews were administered three months after the end of the summer recreation season, it is possible that the respondents have underestimated their true expenditures. In view of these difficulties, we have decided throughout this chapter to replace price with distance, which is easily computed for all sites. This is a quite common practice in recreation studies and is justified if travel and access costs are proportional to distance. That this might be so is suggested by the following regression of access costs, as reported by respondents, on distance (in miles):

$$\underset{(1.06)}{\text{Price} = 0.0949} + \underset{(4.73)}{0.04086}\ \text{Distance} \qquad R^2 = .012 \quad F = 22.35$$

(1214 observations)

Table VII-5
Occurrences of Zero Expenditures for Site Visits

Site	# of Mentions	# of Mentions with Zero Expenditures for Access	On-site
1	21	20	11
2	45	35	32
3	98	56	45
4	112	94	54
5	9	8	6
6	15	13	12
7	14	12	10
8	30	29	20
9	7	7	5
10	11	11	8
11	9	9	3
12	11	10	4
13	11	7	7
14	4	0	1
15	30	30	28
16	115	81	51
17	51	27	29
18	74	30	41
19	43	29	30
20	23	14	15
21	15	10	9
22	34	34	25
23	14	13	9
24	17	16	15
25	20	20	18
26	48	45	39
27	8	5	7
28	22	9	18
29	2	1	2
30	24	24	22
31	43	42	41
32	18	9	8
33	10	3	6
34	4	4	2
35	9	9	8
36	4	3	3
37	4	4	4
38	27	24	20
39	24	17	23
40	18	11	5
41	11	11	8
42	49		
43	6	5	1
All Sites	1164	241	408

NOTE: This table excludes those respondents who mentioned a site for the purpose of rating its characteristics but did not actually visit it.

2. Some Determinants of Recreation Activity

Although the following sections present demand functions for individual sites, it is interesting to consider how the total number of sites visited or the total number of visits to all sites per household is affected by various socio-economic and demographic factors, Some regressions with thse dependent variables are shown in Table VII-6. The first equations deal with household income and structure. KIDS is the number of persons aged 17 and under in the respondents' household; PEOPLE is the total number of persons of all ages in the household. We might expect that the number of children in the household would have a stronger effect on the scope of the household's beach recreation activity than the total size of the household. The opposite appears to be the case:* in no case was the slope coefficient significantly different from zero for KIDS. Also, it appears that the household income has no influence on the total number of visits to all sites by household (although it does affect the total number of sites visited--richer families are likely to visit more sites than poor families). However, the relationship is fairly weak and is complicated by the collinearity between household income and size.**

The next two regressions deal with racial differences in recreation activity. IRISH is dummy variable taking the value 1 if the respondent described himself as having an Irish background. ITALIAN is a dummy variable for respondents with an Italian background and OTHER CAUCASIAN is a dummy variable for other Caucasian backgrounds. Thus the slope coefficients represent differential effects relative to respondents from minority groups--American Indian, Asian-American, Black and Spanish Surname. In the regressions of both numbers of visits to all sites and number of sites visited

---

*Similar results were obtained when we used a dummy variable taking the value 1 if there were children and 0 if there were none, in place of the continuous variable KIDS.

**In these regressions we have replaced missing household income values with the sample mean, $14,137. This is the so-called zero-order regression method--see Afifi and Elashoff [1]; in the present context it produces unbiased but inefficient estimates.

Table VII-6

Total Site Visitation as a Function of Selected Socioeconomic Characteristics

(462 observations)

$$\# \text{ VISITS} = \underset{(.37)}{8.309} + \underset{(.30)}{0.7117} \ln(\text{INC}) + \underset{(2.08)}{1.4317} \text{ PEOPLE} \qquad R^2=.01 \quad F=2.41$$

$$\# \text{ VISITS} = \underset{(3.86)}{15.012} - \underset{(.08)}{1373.79}/\text{INC} + \underset{(2.13)}{1.4628} \text{ PEOPLE} \qquad R^2=.01 \quad F=2.37$$

$$\# \text{ SITES} = \underset{(.08)}{0.116} + \underset{(1.58)}{0.2428} \ln(\text{INC}) + \underset{(.89)}{0.1736} \text{ KIDS} \qquad R^2=.008 \quad F=1.74$$

$$\# \text{ SITES} = \underset{(13.26)}{2.611} - \underset{(1.82)}{2099.74}/\text{INC} + \underset{(.92)}{0.1783} \text{ KIDS} \qquad R^2=.009 \quad F=2.14$$

$$\# \text{ SITES} = \underset{(.27)}{0.39} + \underset{(1.12)}{0.1746} \ln(\text{INC}) + \underset{(2.55)}{0.1148} \text{ PEOPLE} \qquad R^2=.02 \quad F=4.59$$

$$\# \text{ SITES} = \underset{(8.67)}{2.203} - \underset{(1.43)}{1654.43}/\text{INC} + \underset{(2.55)}{0.1142} \text{ PEOPLE} \qquad R^2=.022 \quad F=4.98$$

$$\text{INC} = \underset{(13.5)}{11266.4} + \underset{(4.09)}{732.145} \text{ PEOPLE} \qquad R^2=.035 \quad F=16.75$$

$$\# \text{ VISITS} = \underset{(3.20)}{13.94} + \underset{(3.01)}{16.02} \text{ IRISH} + \underset{(0.95)}{5.2} \text{ ITALIAN} + \underset{(1.13)}{5.4} \text{ OTHER CAUCASIAN} \qquad R^2=.026 \quad F=4.07$$

$$\# \text{ SITES} = \underset{(6.89)}{1.98} + \underset{(2.7)}{0.95} \text{ IRISH} + \underset{(2.13)}{0.77} \text{ ITALIAN} + \underset{(1.22)}{0.38} \text{ OTHER CAUCASIAN} \qquad R^2=.022 \quad F=3.38$$

$$\# \text{ VISITS} = \underset{(5.33)}{10.30} + \underset{(0.85)}{3.213} \text{ AUTO OWNERSHIP} \qquad R^2=.002 \quad F=.73$$

$$\# \text{ SITES} = \underset{(9.51)}{2.14} + \underset{(1.81)}{0.448} \text{ AUTO OWNERSHIP} \qquad R^2=.007 \quad F=3.29$$

$$\# \text{ VISITS} = \underset{(4.78)}{19.46} + \underset{(0.39)}{0.527} \text{ DAYS WORKED PER WEEK} \qquad R^2=.000 \quad F=.16$$

$$\# \text{ SITES} = \underset{(7.4)}{1.973} + \underset{(2.15)}{0.188} \text{ DAYS WORKED PER WEEK} \qquad R^2=.01 \quad F=4.6$$

the hypothesis that the slope coefficients are all zero--if there is no difference in the recreation behavior of minority and other groups--is rejected at the .05 level. However, in the first regression, it is clear that only the Irish have a significantly different recreation behavior--on average they make 16 more visits per household--while Italian and other Caucasian respondents have the same behavior as minority group respondents. In the case of the number of sites visited, both Irish and Italian, but not other Caucasians, have a significantly different behavior from minority groups; moreover, the hypothesis that Irish and Italian respondents have the <u>same</u> behavior cannot be rejected at the .05 level.

The remaining regressions show that automobile ownership has some effect on the number of sites visited, but not on the total of visits to all sites: and also that the length of the working week has a similar effect. However, the sign of the relationship is the opposite of what we might expect--it appears that longer working weeks lead to a larger number of sites being visited. In some regressions not reported here, we found no relationship between the length of paid vacation and the total number of visits to all sites or the total number of sites visited. This is not surprising since our data pertains to day trips and we might expect vacation length to influence more extended trips but not day trips.

3. Abstract Site Demand Functions

The demand functions presented in this section differ from the demand functions to be discussed in the next section in two ways. Firstly, the demand functions presented in this section contain only own price and quality variables. Secondly, they are not derived from an explicity utility function.* On the other hand, the demand functions in this section differ from those estimated by Clawson and Knetch [5], and those who have copied their methodology in that instead of estimating separate demand functions for each site or for groups of sites and included site quality explicity as an explanatory variable. The demand functions thus resemble the "abstract mode" demand functions pioneered in transportation economics by Quandt and Baumol [9]. The functions which we estimate have the following form

$$V_{it} = f[d_{it}, Z_i, C_{it}, Y_t] \qquad \text{... (1)}$$

where V_{it} is the number of visits to a site i by an individual t, d_{it} is the distance traveled (a proxy for price) for individual t in visiting site i, Z_i is a vector of "objective" characteristics of site i, C_{it} is a vector of characteristics of site i as perceived by individual t, and y_t is a fector of characteristics pertaining to individual t, such as household income and composition.

At this point we must deal with the question of zero visitation rates. As Table VII-4 indicates, nobody in our sample visits all of the possible sites and indeed, most people visit very few of them. We re-

*In Section 3 of Chapter 3 we suggested a specific utility function which would lead to demand functions containing only own price and quality variables--see equation (13) of Chapter 3. However, as we pointed out, these particular demand functions require a form of constrained estimation which would be very burdensome computationally, and we have not attempted to estimate them.

marked in Chapter III that the problem of zero visitation rates can be incorporated into stochastic choice system demand models, but it would be prohibitively expensive to apply such a model when there are so many alternative sites. It is relatively easier to deal with this phenomenon in the context of the ad-hoc demand functions represented by (1). Since there are 4627 respondents in our sample and 29 sites (at least), (V_{it}) would be a vector with 13,543 (= 467 x 29) rows. 912 elements of (V_{it}) would be non-zero--this is the number of mentions corresponding to sites 1-29, as listed in Table VII-1--and the remainder would be zero. The obvious estimation method would be Tobit analysis.* Unfortunately, however, the data sets involved are too large to be handled by the conventional Tobit programs. The alternative is a two-step procedure suggested by Goldberger [6], in which the analysis is broken down into two issues.** The first issue is what determines whether a given individual visits a given site at all. We can think of the dependent variable, V_{it}, as being a dummy variable which takes the value 1 if individual t makes at least one visit to site i, and the value zero otherwise. Thus (V_{it}) is a 13543 x 1 vector of 1's and 0's. The second issue is: given that an individual visits a site, what determines how many times he visits it? In this case, the analysis is restricted to the subject of cases where visits are actually made, and the dependent variable, V_{it}, is a 912 x 1 vector containing the (non-zero) numbers of visits by each household to each site.

The two-stage procedure does not necessarily produce the same coefficient estimates as the theoretically preferable Tobit analysis, but is is the best alternative available. Moreover, as Goldberger [6] points out, it is somewhat more flexible than the Tobit procedure because it allows us to specify different,sets of regressors in the two stages of the estimation. Thus the factors which determine the probability of an individual's making any visit to a site need not be the same as those which determine how many

*See, e.g., Goldberger [6].

**Goldberger [6].

visits he makes to those sites which he does visit. We intend to exploit this opportunity; indeed it is necessary for us to do so because, as noted in Section 1, subjective site ratings are generally available only for those sites which respondents actually visited. Thus these variables can be included in the second, but not the first stage regression. Moreover, in our opinion, certain socio-economic variables such as household income and size are not likely to influence whether an individual visits a random site, although they are likely to influence how many visits an individual makes to a site which he does visit?* Therefore, we propose to exclude these two variables from the first stage regressions.

The first-stage regressions, although computationally more convenient than Tobit analysis, are by no means problem free. The dependent variable in those regressions is a dummy variable and OLS is not a natural estimation method in these circumstances. The normal practice is to use maximum likelihood estimates based on some specification of the random process which generates the 1's and 0's,, the most common specifications being the Probit and the Logit models. The two models are quite similar but, Since the latter is more convenient for reasons to be explained below, we adopt it here. The idea behind Logit (and Probit) analysis is similar to the idea behind the discrete dependent variable model presented in Chapter III. We assume that there is an underlying unobserved continuous variable W given by

$$W = \alpha + \Sigma x_j \beta_j + \tilde{u} \qquad \ldots (2)$$

and the observed dichotomous variable V is generated from W by the rule

*This statement may not be strictly true in the light of the results reported in Section 2. An alternative statement, which may be more acceptable, is that the influence of household income and size on the probability that an arbitrary individual visits an arbitrary site is less interesting than the influence of these variables on the number of visits made by an individual to those sites which he does visit.

$$V = \begin{cases} 1 & \text{if } W>0 \\ 0 & \text{if } W<0 \end{cases} \qquad \ldots \text{(3)}$$

Thus if H ($\cdot$) is the cumulative distribution function of the random variable $\tilde{u}$, we have:

$$P = \text{Prob}[V-1] - H[-\alpha - \Sigma x_j \beta_j]$$
$$(1-P) = \text{Prob}[V-0] = 1-H[-\alpha - \Sigma x_j \beta_j]$$

If $\tilde{u}$ is assumed to be normally distributed, we have the Probit Model; if $\tilde{u}$ is assumed to follow the logistic distrubtion, we have the Logit model. In the latter case we observe that

$$\log \frac{P}{1-P} = \alpha + \Sigma x_j \beta_j$$

and

$$P = \frac{1}{1 + e^{-\alpha - \Sigma x_j \beta_j}}$$

For either model the likelihood function is

$$\mathcal{L} = \prod_{V_{it}=1} H[-\alpha-\Sigma x_j \beta_j] \prod_{V_{it}=0} (1-H[-\alpha-\Sigma x_j \beta_j]) \quad \ldots \text{(4)}$$

It would be possible to obtain maximum likelihood estimates of the coefficients ($\Sigma x_j \beta_j$) on the basis of (4) but, given the size of our data set, this would be very expensive. Instead we shall avail ourselves of a much simpler computational procedure suggested recently by Haggerstrom [7] on the basis of work by Halperin, Blackwolder and Verter [8]. The latter authors show that maximum likelihood estimates of the parameters of the Logit model in practice are very close to the coefficient estimates obtained by discriminant analysis. Haggerstrom points out that discriminant analysis coefficients can be obtained from a relatively simple transformation of ordinary least squares regression coefficients using a dummy-dependent variable. Thus, while OLS by itself is not an appropriate technique for handling dummy dependent variables, the OLS coefficient estimates when suitably transformed provide a good approximation to the maximum likelihood estimates of the Logit coefficients, and the OLS t and F statistics may reasonably be used to test hypotheses about the Logit coefficients. It should be noted that, although the predicted values of the dependent variable obtained using OLS are not constrained to lie between 0 and 1, the predicted values of the dependent variable obtained from the transformed OLS coefficients do satisfy this constraint. Haggerstrom shows that, if (α, β_j) are the OLS coefficient estimates and $(\hat{\alpha}, \hat{\beta}_j)$ the discriminant analysis coefficient estimates the required transformation is:

$$\hat{\beta}_j = C\beta_j$$

$$\hat{\alpha} = \log (P_1/P_2) + C[\alpha - \tfrac{1}{2}] + \frac{n}{2}[n_1^{-1} - n_2^{-1}]$$

where c = n/SSR, SSR being the sum of squared residuals from the OLS regression n_1 is the number of cases in which the dependent variable takes the value 1 (i.e. 912), $n_2 = n - n = 12{,}486$, $P_1 = n_1/n$, and $P_2 = n_2/n$.

For the reasons mentioned above, we decided that the most important regressor variables for the first stage analysis were the distance of individual t from the site i and some measures of water quality at site i. On the basis of the regression analysis of willingness to pay and the accuracy of subjective perception of water quality parameters reported elsewhere, we decided to confine our analysis to three parameters--color, coliform bacteria counts and phosphorous content. When we came to implement the OLS regression of a dummy variable for site visitation we found that, even using OLS, the data set exceeded the capacity of the programs available to us, so we restricted ourselves to no more than two regressions and truncated the data set at 11,000 observations. The results of these regressions are shown in Table VII-7. The regression coefficients have the signs which we would expect and are significantly different from zero: the greater the distance and the more polluted a site (in terms of color, coliform bacteria or phosphorus) the lower the

Table VII-7

Probability of Site Visitation -- Logit Model

Variable	OLS Estimate	Discriminant Estimate	OLS Estimate	Discriminant Estimate	OLS Estimate	Discriminant Estimate
CONSTANT	0.1682 (25.6)	-1.245	0.1094 (22.73)	-2.0437	0.1944 (27.79)	-0.392
DISTANCE	-0.00433 (11.5)	-0.06543	-0.003 (8.18)	-0.04465	-0.00533 (13.73)	-0.08129
PHOSPHORUS	-0.7332 (13.95)	-11.0799				
COLI			-0.00000315 (4.97)	-0.0000469		
COLOR					-0.00803 (17.26)	-0.1232
R^2	.022		.007		.031	
F	120.42		39.12		176.38	
SSR	727.91		739.12		721.25	
n/SSR		15.11		14.88		15.25

$n = 11{,}000;\ n_1 = 803;\ n_2 = 10{,}197$

$\log(P_1/P_2) + \frac{n}{2}\left[\frac{1}{n_1} - \frac{1}{n_2}\right] = \quad 3.768$

probability that a respondent visits it. The impact of objective water quality conditions on the probability that a site is visited at least once is unambiguously established by these results.

However, when we come to the second stage regressions--the OLS regression of the number of visits by members of a respondent's household to each site which it visits--we reach a rather different conclusion. Tables VII-8 and VII-9* presents the results of several regressions of this variable on various sets of regressors including alternatively subjective water quality ratings and objective measures of water quality. The other variables are distance from site (DIST), household income** and size (INC, PEOPLE) and a dummy variable, ACCESS, which takes the value 1 if the site is accessible by public transportation and the value 0 otherwise.*** Several results stand out in these regressions. DIST always has a significant negative coefficient and although the coefficient of INC is unstable in sign and frequently insignificant--at least partly because of the colinearity with PEOPLE--in the preferred equations it is positive and fairly significant. As we might expect, household size and accessibility to public transport always have a positive effect on the number of visits to a site although these slope coefficients are not always significant.

The most important findings concern the relative performance of subjective and objective measures of water quality as explanatory variables. Subjective water quality rating always has a significant positive coefficient--respondents make more visits to a site which they consider to be of higher quality. This is not a surprising conclusion, although

*There are 819 rather than 912 observations because 93 site cards contain no water or beach quality ratings

** As with the regressions presented in section 2, we have replaced missing income values with the mean income of $14,317.

*** In Tables VII-7 and VII-8 an asterisk marks the preferred equation. The choice between functional forms is based on the Box & Cox [3] maximum likelihood criterion.

Table VII-8

Abstract Site Demand Functions with Subjective Quality Ratings

(819 observations)

VISITS = 4.226 + 7903.46/INC + 0.8649 RWQUAL - 0.3775 DIST
(2.76) (1.53) (2.68) (5.73)

+ 2.2461 ACCESS - 0.3804 PEOPLE
(2.63) (2.13)
R^2=.076 F=13.32

VISITS = 5.431 - .00000675 INC + 0.8567 RWQUAL - 0.3889 DIST
(3.79) (.14) (2.65) (5.89)

+ 2.2323 ACCESS + 0.3346 PEOPLE R^2=.073 F=12.82
(2.61) (1.86)

VISITS = 11.441 - 0.671n(INC) + 0.8567 RWQUAL - 0.3811 DIST
(1.84) (1.00) (2.66) (5.77)

+ 2.2376. ACCESS + 0.3665 PEOPLE R^2=.074 F=13.033
(5.76) (2.60)

1n(VISITS) = 1.363 + 512.653/INC + 0.0745 RWQUAL - 0.0464 DIST
(10.39) (1.16) (2.7) (8.23)

+ 0.088. ACCESS + 0.0172 PEOPLE R^2=.103 F=18.573
(1.2) (1.12)

1n(VISITS) = 1.375 + 512.887/INC + 0-0776 RWQUAL - 0.0077 RBQUAL
(9.93) (1.16) (2.61) (.27)

- 0.0463 DIST + 0.0875 ACCESS + 0.0175 PEOPLE R^2=.103 F=15.473
(8.22) (1.20) (1.14)

*1n(VISITS) = 13.607 + 0.0000072 INC + 0.0759 RWQUAL - 0.0485 DIST
(11.13) (1.78) (2.76) (8.62)

+ 0.0861 ACCESS + 0.00856 PEOPLE R^2=.105 F=18.98
(1.18) (5.56)

1n(VISITS) = 1.409 + 0.003041n(INC) + 0.0741 RWQUAL - 0.0472 DIST
(2.65) (.05) (2.69) (8.36)

+ 0.087 ACCESS + 0.0138 PEOPLE R^2=.101 F=18.275
(1.19) (0.90)

Table VII-9

Abstract Site Demand Functions with Objective Quality Variables for 29 Sites

$$\text{VISITS} = \underset{(3.92)}{5.83} + \underset{(1.48)}{7693.71}/\text{INC} + \underset{(0.26)}{0.0000344}\ \text{COLI} + \underset{(0.57)}{0.0526}\ \text{COLOR}$$

$$- \underset{(4.67)}{0.3025}\ \text{DIST} + \underset{(2.21)}{1.9361}\ \text{ACCESS} + \underset{(2.38)}{0.426}\ \text{PEOPLE} \qquad R^2=.068 \quad F=9.901$$

$$\text{VISITS} = \underset{(3.69)}{5.525} + \underset{(1.37)}{7110.52}/\text{INC} - \underset{(0.44)}{0.000054}\ \text{COLI} + \underset{(0.4)}{0.044}\ \text{COLOR}$$

$$+ \underset{(1.68)}{24.217}\ \text{PHOSPHORUS} - \underset{(4.52)}{0.2935}\ \text{DIST} + \underset{(2.05)}{1.8}\ \text{ACCESS}$$

$$+ \underset{(2.14)}{0.387}\ \text{PEOPLE} \qquad R^2=.071 \quad F=8.907$$

$$\text{VISITS} = \underset{(5.07)}{7.023} - \underset{(0.23)}{0.00001105}\ \text{INC} + \underset{(0.26)}{0.0000352}\ \text{COLI} + \underset{(0.57)}{0.0531}\ \text{COLOR}$$

$$- \underset{(4.84)}{0.3133}\ \text{DIST} + \underset{(2.20)}{1.9239}\ \text{ACCESS} + \underset{(2.14)}{0.3843}\ \text{PEOPLE} \qquad R^2=.066 \quad F=9.52$$

$$\ln(\text{VISITS}) = \underset{(12.12)}{1.54} + \underset{(1.12)}{495.99}/\text{INC} + \underset{(1.29)}{0.0000146}\ \text{COLI} - \underset{(0.4)}{0.00318}\ \text{COLOR}$$

$$- \underset{(7.36)}{0.0408}\ \text{DIST} + \underset{(0.69)}{0.0514}\ \text{ACCESS} + \underset{(1.45)}{0.0222}\ \text{PEOPLE} \qquad R^2=0.96 \quad F=14.424$$

$$\ln(\text{VISITS}) = \underset{(12.02)}{1.54} + \underset{(1.11)}{494.95}/\text{INC} + \underset{(1.17)}{0.0000144}\ \text{COLI} - \underset{(0.36)}{0.00335}\ \text{COLOR}$$

$$+ \underset{(0.03)}{0.043.}\ \text{PHOSPHORUS} - \underset{(7.33)}{0.0407}\ \text{DIST} + \underset{(0.68)}{0.0152}\ \text{ACCESS}$$

$$+ \underset{(1.43)}{0.0221}\ \text{PEOPLE} \qquad R^2=.096 \quad F=12.35$$

$$*\ln(\text{VISITS}) = \underset{(13.05)}{1.541} + \underset{(1.65)}{0.00000672}\ \text{INC} + \underset{(1.27)}{0.0000143}\ \text{COLI} - \underset{(0.35)}{0.00279}\ \text{COLOR}$$

$$+ \underset{(7.72)}{0.426.}\ \text{DIST} + \underset{(0.66)}{0.0494}\ \text{ACCESS} + \underset{(0.91)}{0.014}\ \text{PEOPLE} \qquad R^2=.098 \quad F=14.7$$

$$\ln(\text{VISITS}) = \underset{(2.98)}{1.59} + \underset{(0.04)}{0.0021}\ \ln(\text{INC}) + \underset{(1.3)}{0.0000146}\ \text{COLI} - \underset{(0.4)}{0.0031}\ \text{COLOR}$$

$$- \underset{(7.5)}{0.0416}\ \text{DIST} + \underset{(0.68)}{0.0505}\ \text{ACCESS} + \underset{(1.22)}{0.0188}\ \text{PEOPLE} \qquad R^2=.095 \quad F=14.2$$

the direction of causation is ambiguous. It might be best to regard site ratings as jointly endogenous variables together with site visitation rates, the true exogenous variables being the objective measures of site quality. However, there is very little relationship between objective measures of site quality and the frequency with which a site is visited. The coefficients of COLOR, COLI BACT and PHOS are usually insignificant and frequently of the "wrong" sign. The data provides little evidence that objectively better sites are visited more frequently, other things being equal.

Thus, we may conclude that if a site has a better water quality there is a higher probability that a household taken at random will visit it at least once but, given that the household does visit the site, there is little reason to believe that the site is visited more frequently than other sites of lower water quality. On the other hand, households make more visits to sites which they believe to be of a higher quality--or perhaps the converse is true: households believe that the sites which they visit often are better than those which they visit rarely. This discrepancy is similar to that observed in the analysis of willingness-to-pay; households were willing to pay more for sites which they believed to be of a higher quality, but not necessarily for sites wihich objectively had a higher quality. It is consistent with our finding in Chapter 5 that subjective site rating match up with objective site conditions only imperfectly.

4. System Demand Functions

Chapter 3 suggested the following model for deriving site demand functions based on p characteristics Z_{ij}:

$$U = \Sigma b_i \log(V_j - c_i) \qquad \ldots \text{(1a)}$$

$$c_i = W_{io} + \sum_{k=1}^{p} W_k Z_{ik} \qquad \ldots \text{(1b)}$$

The demand functions obtained from this utility model are:

$$V_{it} = c_i + \frac{b_i}{\sum_{j=1}^{n} b_j} \frac{1}{P_{it}} \left[Y_t - \sum_{j=1}^{n} P_{jt} c_j\right] \quad i=1\ldots n \ \ldots \text{(2)}$$

where V_{it} is the number of visits to site i by individual t. The standard practice in consumer demand theory is to normalize the b_i's so that $\Sigma b_i = 1$, in which case (2) can also be written in expenditure form as

$$P_{it} \cdot V_{it} = P_{it} c_i + b_i \left[Y_t - \Sigma P_{jt} c_j\right] \qquad \ldots \text{(3)}$$

This function is nonlinear in the parameters b_i and c_i (or, equivalently in the parameters b_i and W_{io}, W_k). Two alternative estimation procedures are available: a maximum likelihood estimation procedure due to Parks [10] and less sophisticated iterative two-part procedure due to Stone [12]. Because of its computational simplicity, we shall follow Stone's procedure here. This procedure is based on the fact that, for a given set of values of the parameters b_i, equations (2) and (3) are linear functions of c_i (or, equivalently, of W_{io} and W_k), while for a given set of values of the parameter c_i, these equations are linear functions of b_i. Stone's method is to iterate between OLS estimates of b_i, for given values of c_i, and OLS estimates of c_i, for given values of b_i.

At this point we have to face the fact, hitherto neglected, that we are actually dealing with a subset of commodities--namely, expenditures on recreation sites--rather than with the whole set of consumption items. This raises the question of whether the theory developed for the latter situation can be applied here. The answer is that the general theory does carry over to the case of a subset of commodities if the consumer's utility function is assumed to be appropriately separable. There are various concepts of separability which we might invole; without going into detail, we may state that an underlying idea of these concepts is that the marginal rate of substitution between any pair of recreation sites should be independent of the consumer's level of consumption of any other commodity besides recreation sites.* This is a strong requirement, but not an entirely unreasonable one. If it is accepted, and if the relevant portion of the consumer's utility function dealing with the utility from beach recreation is given by (1), then the site demand functions are indeed given by (2) or (3), with one change. Site depand depends on the prices of the n sites and on the total expenditure on beach recreation, rather than income. Thus, the variable, Y, in (2) or (3) must be taken as standing for total expenditure on water-oriented recreation. This variable is then endogenous to the consumer's choice process, and is, therefore, a function of the prices of both recreation sites and (in general) all the other commodities as well as income. Instead of trying to model the determinants of recreation expenditure explicitly, we shall employ the assumption commonly used in Engle curve analysis

*See, for example, Pollak [11].

that there is relatively little variation in the prices of non-recreation goods faced by our sample households; hence, we may postulate some simple relationship between expenditure on beach recreation (Y) and income, such as

$$Y = d_o + d_i INC \qquad \text{... (4a)}$$

or

$$Y = d_o + d_1 \ln(INC) \qquad \text{... (4b)}$$

If we substitute (4a) or (4b) into (2) or (3) we have a fully specified system of demand equations for recreation sites, under the separability assumption.

There are still some complications due to the fact that, for the reasons outlined in Section 1, we do not have good price data. Because of this deficiency, we have chosen to use distance as a proxy for price and, as we observed in the previous section, this seems to be a good substitute. However, in the context of system demand models, this substitution causes some problems because it means that the "adding-up condition" no longer applies--i.e., it is no longer true that for each individual, the sum of the left-hand side variables in Equation (3) over all sites is exactly equal to Y, the total expenditure on water-oriented recreation. The adding-up condition in practice has an important role in the estimation of (2) or (3) both with the maximum likelihood procedure and with Stone's method. In the latter case it helps to ensure that $\Sigma b_i = 1$ without the need for constrained estimation techniques. Without this assumption, therefore, we must either use constrained OLS estimation, which is computationally difficult or simplify the model further. We have chosen the latter alternative. Specifically, we have assumed that

$$b_i = b \qquad i = 1 \ldots n \qquad \text{... (5)}$$

and, without any loss of generality, we have taken b=1. Accordingly, the term $(b_i/\Sigma b_j)$ in (2) is replaced by (1/n), n being the number of sites. Since we have in effect suppressed b_i as a parameter, the only parameters to be estimated are the c_i's (i.e., W_{io}, W_k); as we noted above, with the values of b_i known, equations (2) or (3) are linear in the latter variables and a single-stage OLS estimation may be applied. We have, thus, removed the need for iterating on the coefficient estimates, thus greatly reducing the computational difficulty.

The model which we propose to estimate is given by (2), (1b), (4) and (5). We have chosen to use as site characteristics COLOR and COLIFORM; thus, there are 33 coefficients to be estimated: 29 W_{io}'s--one for each site; W_1, the coefficient of COLOR; W_2, the coefficient of COLI; and the parameters d_o and d_1 in (4). We have 912 observations from which to estimate these coefficients, corresponding to the site cards with non-zero visits. Assuming that we share the specification (4a), the actual estimating equations are:

$$P_iV_i = W_{io}P_i\left(\frac{n-1}{n}\right) - \sum_{j\neq i} W_{jo}\frac{P_j}{n} + W_1\left\{Z_{1i}P_i\left(\frac{n-1}{n}\right) + \sum_{j\neq i} Z_{1j}P_{j/n}\right\}$$

$$+ W_2\left\{Z_{2i}P_i\left(\frac{n-1}{n}\right) + \sum_{j\neq i} Z_{2i}P_{j/n}\right\}$$

$$+ \frac{d_o}{n} + \frac{d_1}{n}\,\text{INC} \qquad i=1\ldots n \qquad \ldots (6)$$

Unfortunately, despite several attempts to model (6), we were unable to do so. The reason was that the data were highly collinear leading to a nearly singular cross-product matrix which could not be inverted. One possible solution may be to group neighboring sites of similar quality so that there is a smaller set of sites differing more in their locations. This would cause the matrix of price (distance) variables to be less collinear and simultaneously reduce the number of parameters to be estimated. Also, it is possible that maximum likelihood estimation of a less specialized version of the model might prove to be more successful. There is ample scope for further research on the specification and estimation of the model, but this was beyond the scope of this project.

5. Benefit Calculation

The only rigorous method to obtain empirical measures of willingness-to-pay for changes in recreation site quality is to estimate a set of demand functions which can be shown to derive from a specific utility function and, using the coefficient estimates, to calculate the resulting change in the area under the compensated function. If the utility function is that given by (1), the corresponding formula for the consumer's surplus associated with a change in site quality is given by Formula (1) in Chapter 3, with the c_i terms replaced by equations (1b) above. Since we are not presently able to estimate this demand model, we are unable to apply this methodology to calculate the benefits of changes in water quality.

We are forced, instead, to rely on the abstract site demand functions described in Section 3. Since these demand functions are not derivable from an explicit utility function, there is no basis for calculating measures of consumer surplus. All that we can do with these demand functions is to predict the impact of water quality changes in site visitation. The only solution is to use some ad hoc metric such as the Principals and Standards estimate that one visitor day is "worth" $.75-$2.50; alternatively, we could value visits at the average willingness-to-pay plus transportation costs as expressed by the respondents to our questionnaire. As an illustration of this procedure, suppose that the coliform bacteria count at a site declines from the average (2000) to the minimum across the samples, 100. Assuming that an individual lives five miles from the site; using the coefficients in the fourth column of Table VII-7, we calculate that the probability that the individual will visit the site changes from:

$$P = \frac{1}{1+e^{2.364}} = 0.086 \text{ (COLI=2000)}$$

or

$$P = \frac{1}{1+e^{2.2716}} = 0.094 \text{ (COLI =100)}$$

If we assume that the individual makes eight visits to a chosen site, and this number is not affected by the change in water quality, the expected visitation of the site changes from 0.69 visits (=0.086x8) to 0.75 visits (=0.094x8). Valuing each visit at $2.50 per person and assuming that there are four persons in the group, the dollar value of the change in water quality for this household is 5.64 (=$2.50 x 0.064 x 4). This would equal something less than $400,000 for the whole SMSA, integrating over distance, or $410,000 if the site was five miles from the bulk of the population. This is no doubt a substantial underestimate of the total benefit of the hypothetical coliform reduction since, as Chapter II explains, consumer's surplus has been ignored. The point of this example is principally to illustrate how the abstract site model can be used.

6. Conclusions

This chapter provides interesting additional evidence for some of the points argued elsewhere in the report. First, persons with large families or families with higher incomes tended to visit our sample beaches more frequently than other families. Family ethnic background also appears to influence recreation behavior.

Second, substitution between sites is a significant aspect of recreation behavior in the Boston sample of households and sites. Most respondents visited two or more sites during the summer. Under direct questioning, most cited inter-site substitution as their most likely response to a change in water quality at their favorite beach. Any-where proximal sites are close substitutes, perhaps most urban areas, inter-site substitution is likely to be an important phenomenon. Thus, single site demand models are not altogether appropriate for either demand forecasting or benefit estimation. We specified a system demand model to account explicitly for this behavior, but were not able to complete its estimation with the resources available to us. This is a fruitful area for further research.

Finally, poor water quality at a site appears to reduce the pro-bability that a randomly selected household will visit the site at all, but does not influence the number of visits to the site given that it is visited at least once. Hence, water quality changes impact recreation behavior principally through inter-site substitutions; this reinforces the need for systems demand models. On the other hand, higher perceived water quality is significantly associated with more visits, but the dir-ection of causation is by no means evident. Again we must conclude that while subjective ratings of water quality match only poorly objective measures, Bostonians seem to value maintaining and improving the area's waters for recreational uses.

CITED REFERENCES

1. A.A. Afifi and R.M. Elashoff, "Missing Observations in Multivariate Statistics II. Point Estimation in Simple Linear Regression," Journal of the American Statistical Association (March 1967): 10-29.

2. M.R. Anderberg, Cluster Analysis for Applications, Academic Press, 1973.

3. G.E.P. Box and D.R. Cox, "An Analysis of Transformations," Journal of Royal Statistical Association Series B (1964): 211-252.

4. C.J. Cicchetti, A.C. Fisher, V. Kerry Smith, "An Economic Evaluation of a Generalized Consumer Surplus: The Mineral King Controversy," unpublished paper, Natural Environments Program, Resources for the Future, 1973.

5. M. Clawson & J. Knetsch, Economics of Outdoor Recreation, Baltimore: Johns Hopkins University Press, 1966.

6. A. Goldberger, Econometric Theory, New York: John Wiley & Sons, 1966.

7. Haggestrom, Notes on Discriminant Analysis, Logistic Regression, Rand Memorandum, dated 4/3/73.

8. M. Halperin, W.C. Blackwalder & J.C. Verter, "Estimation of the Multivariate Logistic Function: A Comparison of the Discriminant Function and ML Approaches," Journal of Chronic Diseases (1971): 125-158.

9. R.E. Quandt & W.J. Baumof, "The Demand for Abstract Transport Modes: Theory & Measurement," Journal of Regional Science (1966): 13-26.

10. R.W. Parks & A.P. Barten, "A Cross-Country Comparison of the Effects of Prices, Income & Population Composition on Consumption' Patterns," Economic Journal (September 1973).

11. Pollak, Robert, "Subindexes in the Cost of Living Index," International Economic Review (February 1975): 135-150.

12. Stone, The Measurement of Consumers' Expenditure and Behavior in the U.K., 1820-1938, Volume 1, Cambridge University Press, 1953.